TRUNCATED RANDOM MEASURES

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JANUARY 2018

FINAL TECHNICAL REPORT

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ **AIR FORCE MATERIEL COMMAND**   ■   **UNITED STATES AIR FORCE**   ■   **ROME, NY 13441**

NOTICE AND SIGNATURE PAGE

AFRL-RI-RS-TR-2018-004 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE CHIEF ENGINEER:

/ S /
CHRISTOPHER J. FLYNN
Work Unit Manager

/ S /
JOHN D. MATYJAS
Technical Advisor, Computing &
  Communications Division
Information Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
JAN 2018	FINAL TECHNICAL REPORT	OCT 2016 – JUL 2017

4. TITLE AND SUBTITLE

TRUNCATED RANDOM MEASURES

5a. CONTRACT NUMBER
FA8750-17-2-0019

5b. GRANT NUMBER
N/A

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Tamara Broderick

5d. PROJECT NUMBER
SAGA

5e. TASK NUMBER
MI

5f. WORK UNIT NUMBER
T1

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Massachusetts Institute of Technology
77 Massachusetts Ave.
Cambridge MA 02139

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory/RITA
525 Brooks Road
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RI

11. SPONSOR/MONITOR'S REPORT NUMBER

AFRL-RI-RS-TR-2018-004

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09

13. SUPPLEMENTARY NOTES

14. ABSTRACT
In many scenarios, we want our model of the world to be able to grow in complexity as we collect more information from the world around us. This growth reflects that we learn more about the world as we acquire more data. And we width to explicitly model both rare events as well as the potential for new events or latent outcomes that we have not yet experienced or collected data on. In this project, we have developed new model representations that enable fast and efficient inference, as well as provided and proved error bounds for certain classes of approximation. The experiments run were on simulated data.

15. SUBJECT TERMS
Bayesian nonparametrics, Markov Chain Monte Carlo, Cumulative density function

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			CHRISTOPHER J. FLYNN
U	U	U	UU	71	19b. TELEPHONE NUMBER *(Include area code)*

Contents

List of Figures

List of Tables

Summary

In many scenarios, we want our model of the world to be able to grow in complexity as we collect more information from the world around us. This growth reflects that we learn more about the world as we acquire more data. And we wish to explicitly model both rare events as well as the potential for new events or latent outcomes that we have not yet experienced or collected data on. In this project, we have developed new model representations that enable fast and efficient inference, as well as provided and proved error bounds for certain classes of approximation. Our experiments below were on simulated data. We started preliminary work on the Innovian Time-Series Anesthesia Dataset but were given access to that data a few weeks before the project concluded and were not able to run our full experiments in that time. We did not create any new data sets as part of this project.

Introduction

In many data sets, we can view the data points as exhibiting a collection of underlying traits. For instance, each document in the *New York Times* might touch on a number of topics or themes, an individual's genetic data might be a product of the populations to which their ancestors belonged, or a user's activity on a social network might be dictated by their varied personal interests. When the traits are not directly observed, a common approach is to model each trait as having some frequency or rate in the broader population [Airoldi et al., 2014]. The inferential goal is to learn these rates as well as whether—and to what extent—each data point exhibits each trait. Since the traits are unknown a priori, their cardinality is also typically unknown.

As a data set grows larger, we can reasonably expect the number of traits to increase as well. In the cases above, for example, we expect to uncover more topics as we read more documents, more ancestral populations as we examine more individuals' genetic data, and more unique interests as we observe more individuals on a social network. *Bayesian nonparametric* (BNP) priors provide a flexible, principled approach to creating models in which the number of exhibited traits is random, can grow without bound, and may be learned as part of the inferential procedure. By generating a countable infinity of potential traits—where any individual data point exhibits only finitely many— these models enable growth in the number of observed traits with the size of the data set.

In practice, however, it is impossible to store a countable infinity of random variables in memory or learn the distribution over a countable infinity of variables in finite time. Conjugate priors and likelihoods have been developed [Orbanz, 2010] that theoretically circumvent the infinite representation altogether and perform exact Bayesian posterior inference [Broderick et al., 2017]. However, these priors and likelihoods are often just a single piece within a more complex generative model, and ultimately an approximate posterior inference scheme such as Markov Chain Monte Carlo (MCMC) or variational Bayes (VB) is required. These approximation schemes often necessitate a full and explicit representation of the latent variables.

One option is to approximate the infinite-dimensional prior with a related finite-dimensional prior: that is, to replace the infinite collection of random traits by a finite subset of "likely" traits. To do so, first enumerate the countable infinity of traits in the full model and write (ψ_k, θ_k) for each paired trait ψ_k (e.g. a topic in a document) and its rate or frequency θ_k. Then the discrete measure $\Theta := \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$ captures the traits/rates in a sequence indexed by k. The (ψ_k, θ_k) pairs are random in the Bayesian model, so Θ is a random measure. In many cases, the distribution of Θ can be defined by specifying a sequence of simple, familiar distributions for the finite-dimensional ψ_k and θ_k, known as a *sequential representation*. Given a sequential representation of Θ, a natural way to choose a subset of traits is to keep the first $K < \infty$ traits and discard the rest, resulting in an approximate measure Θ_K. This approach is called *truncation*. Note that it is also possible to truncate by removing atoms with weights less than a specified threshold [Argiento et al.,

2016, Muliere and Tardella, 1998], though this approach is not as easily incorporated in posterior inference algorithms.

Sequential representations have been shown to exist for *completely random measures* (CRMs) [Kingman, 1967, Ferguson and Klass, 1972], a large class of nonparametric priors that includes such popular models as the beta process [Hjort, 1990, Kim, 1999] and the gamma process [Ferguson and Klass, 1972, Kingman, 1975, Brix, 1999, Titsias, 2008, James, 2013]. Numerous sequential representations of CRMs have been developed in the literature [Ferguson and Klass, 1972, Bondesson, 1982, Rosiński, 1990, 2001, James, 2014, Broderick et al., 2017]. CRM priors are often paired with likelihood processes—such as the Bernoulli process [Thibaux and Jordan, 2007], negative binomial process [Zhou et al., 2012, Broderick et al., 2015], and Poisson likelihood process [Titsias, 2008]. The likelihood process determines how much each trait is expressed by each data point. Sequential representations also exist for *normalized completely random measures* (NCRMs) (sometimes referred to as normalized random measures with independent increments) [Perman et al., 1992, Perman, 1993, James, 2002, Pitman, 2003, Regazzini et al., 2003, Lijoi and Prünster, 2010, James et al., 2009], which provide random distributions over traits, such as the Dirichlet process [Ferguson, 1973, Sethuraman, 1994]. NCRMs are typically paired with a likelihood that assigns each data point to a single trait using the NCRM as a discrete distribution.

Since (N)CRMs have many possible sequential representations, a method is required for determining which to use for the application at hand and, once a representation is selected, for choosing a truncation level. Our main contributions enable the principled selection of both representation and truncation level using approximation error:

1. We provide a comprehensive characterization of the different types of sequential representations for (N)CRMs, filling in many gaps in the literature of sequential representations along the way. We classify these representations into two major groups: *series representations*, which are constructed by transforming a homogeneous Poisson point process; and *superposition representations*, which are the superposition of infinitely many Poisson point processes with finite rate measures. We also introduce two novel sequential representations for (N)CRMs.

2. We provide theoretical guarantees on the approximation error induced when truncating these sequential representations. We give the error as a function of the prior process, the likelihood process, and the level of truncation. While truncation error bounds for (N)CRMs have been studied previously, past work has focused on specific combinations of (N)CRM priors and likelihoods—in particular, the Dirichlet-multinomial [Sethuraman, 1994, Ishwaran and James, 2001, Ishwaran and Zarepour, 2002, Blei and Jordan, 2006], beta-Bernoulli [Paisley et al., 2012, Doshi-Velez et al., 2009], generalized beta-Bernoulli [Roy, 2014], and gamma-Poisson [Roychowdhury and Kulis, 2015] processes. In the current work, we give much more general results for bounding the truncation error.

Our results fill in large gaps in the analysis of truncation error, which is often measured in terms of the L^1 (a.k.a. total variation) distance between the data distributions induced by the full and truncated priors. We provide the first analysis of truncation error for some sequential representations of the beta process with Bernoulli likelihood [Thibaux and Jordan, 2007], for the beta process with negative binomial likelihood [Zhou et al., 2012, Broderick et al., 2015], and for the normalization of the generalized gamma process [Brix, 1999], the σ-stable process, and the

3

generalized inverse gamma [Lijoi et al., 2005, Lijoi and Prünster, 2010] with discrete likelihood. Moreover, even when truncation results already exist in the literature [Ishwaran and James, 2001, Doshi-Velez et al., 2009, Paisley et al., 2012, Roychowdhury and Kulis, 2015], we improve on those error bounds by a factor of two. The reduction arises from our use of the point process machinery of CRMs, circumventing the total variation bound used originally by Ishwaran and James [2001, 2002] upon which most modern truncation analyses are built. We obtain our truncation error guarantees by bounding the probability that data drawn from the full model will use a feature that is not available to the truncated model. Thinking in terms of this probability provides a more intuitive interpretation of our bounds that can be communicated to practitioners and used to guide them in their choice of truncation level.

The remainder of this paper is organized as follows. In Section 4.1, we provide background material on CRMs and establish notation. In our first main theoretical section, Section 4.2, we describe seven different sequential CRM representations, including four series representations and three superposition representations, two of which are novel. Next, we provide a general theoretical analysis of the truncation error for series and superposition representations in Section 4.3. We provide analogous theory for the normalized versions of each representation in Section 4.4 via an infinite extension of the "Gumbel-max trick" [Gumbel, 1954, Maddison et al., 2014]. We determine the complexity of simulating each representation in Section 4.5. In Chapter 5, we summarize our results (Table 1) and provide advice on how to select sequential representations in practice. Proofs for all results developed in this paper are provided in the appendices.

4

Methods, Assumptions, and Procedures

3.1 Background

3.1.1 CRMs and truncation

Consider a Poisson point process on $\mathbb{R}_+ := [0, \infty)$ with rate measure $\nu(\mathrm{d}\theta)$ such that

$$\nu(\mathbb{R}_+) = \infty \qquad \text{and} \qquad \int \min(1, \theta)\nu(\mathrm{d}\theta) < \infty. \qquad (1)$$

Such a process generates a countable infinity of values $(\theta_k)_{k=1}^\infty$, $\theta_k \in \mathbb{R}_+$, having an almost surely finite sum $\sum_{k=1}^\infty \theta_k < \infty$. In a BNP trait model, we interpret each θ_k as the rate or frequency of the k-th trait. Typically, each θ_k is paired with a parameter ψ_k associated with the k-th trait (e.g., a topic in a document or a shared interest on a social network). We assume throughout that $\psi_k \in \Psi$ for some space Ψ and $\psi_k \overset{\text{i.i.d.}}{\sim} G$ for some distribution G. Constructing a measure by placing mass θ_k at atom location ψ_k results in a *completely random measure* (CRM) [Kingman, 1967]. As shorthand, we will write $\mathrm{CRM}(\nu)$ for the completely random measure generated as just described:

$$\Theta := \sum_k \theta_k \delta_{\psi_k} \sim \mathrm{CRM}(\nu).$$

The trait distribution G is left implicit in the notation as it has no effect on our results. Further, the possible fixed-location and deterministic components of a CRM [Kingman, 1967] are not considered here for brevity; these components can be added (assuming they are purely atomic) and the analysis modified without undue effort. The CRM prior on Θ is typically combined with a likelihood that generates trait counts for each data point. Let $h(\cdot \,|\, \theta)$ be a proper probability mass function on $\mathbb{N} \cup \{0\}$ for all θ in the support of ν (though the present work may be easily extended to likelihoods with support in $\mathbb{R}$). Then a collection of conditionally independent observations $X_{1:N} := \{X_n\}_{n=1}^N$ given Θ are distributed according to the *likelihood process* $\mathrm{LP}(h, \Theta)$, i.e.

$$X_n := \sum_k x_{nk} \delta_{\psi_k} \overset{\text{i.i.d.}}{\sim} \mathrm{LP}(h, \Theta),$$

if $x_{nk} \sim h(x \,|\, \theta_k)$ independently across k and i.i.d. across n. The desideratum that each X_n expresses a finite number of traits is encoded by the assumption that

$$\int (1 - h(0 \,|\, \theta))\nu(\mathrm{d}\theta) < \infty. \qquad (2)$$

5

Since the trait counts are typically latent in a full generative model specification, define the observed data $Y_n \mid X_n \overset{\text{indep}}{\sim} f(\cdot \mid X_n)$ for a conditional density f with respect to a measure μ on some space. For instance, if the sequence $(\theta_k)_{k=1}^{\infty}$ represents the topic rates in a document corpus, X_n might capture how many words in document n are generated from each topic and Y_n might be the observed collection of words for that document.

Since the sequence $(\theta_k)_{k=1}^{\infty}$ is countably infinite, it may be difficult to simulate or perform posterior inference in this model. One approximation scheme is to define the *truncation* $\Theta_K := \sum_{k=1}^{K} \theta_k \delta_{\psi_k}$. Since it is finite, the truncation Θ_K can be used for exact simulation or in posterior inference—but some error arises from not using the full CRM Θ. To quantify this error, consider its propagation through the above Bayesian model. Define $Z_{1:N}$ and $W_{1:N}$ for Θ_K analogous to the definitions of $X_{1:N}$ and $Y_{1:N}$ for Θ:

$$Z_n \mid \Theta_K \overset{\text{i.i.d.}}{\sim} \text{LP}(h, \Theta_K), \qquad W_n \mid Z_n \overset{\text{indep}}{\sim} f(\cdot \mid Z_n), \qquad n = 1, \ldots, N.$$

A standard approach to measuring the distance between Θ and Θ_K is to use the L^1 metric between the marginal densities $p_{N,\infty}$ and $p_{N,K}$ (with respect to some measure μ) of the final observations $Y_{1:N}$ and $W_{1:N}$ [Ishwaran and James, 2001, Doshi-Velez et al., 2009, Paisley et al., 2012]:

$$\frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 := \frac{1}{2} \int_{y_{1:N}} |p_{N,\infty}(y_{1:N}) - p_{N,K}(y_{1:N})| \, \mu(\mathrm{d}y_{1:N}).$$

All of our bounds on $\frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1$ are also bounds on the probability that $X_{1:N}$ contains a feature that is not in the truncation Θ_K (cf. Sections 4.3 and 4.4). This interpretation may be easier to digest since it does not depend on the observation model f and is instead framed in terms of the underlying traits the practitioner is trying to estimate.

3.1.2 The gamma-Poisson process

To illustrate the practical application of the theoretical developments in this work, we provide a number of examples throughout involving the *gamma process* [Brix, 1999], denoted $\Gamma\text{P}(\gamma, \lambda, d)$, with discount parameter $d \in [0, 1)$, scale parameter $\lambda > 0$, mass parameter $\gamma > 0$, and rate measure

$$\nu(\mathrm{d}\theta) = \gamma \frac{\lambda^{1-d}}{\Gamma(1-d)} \theta^{-d-1} e^{-\lambda\theta} \mathrm{d}\theta.$$

Setting $d = 0$ yields the undiscounted gamma process [Ferguson and Klass, 1972, Kingman, 1975, Titsias, 2008]. The gamma process is often paired with a Poisson likelihood,

$$h(x \mid \theta) = \frac{\theta^x}{x!} e^{-\theta}.$$

Throughout the present work, we use the rate parametrization of the gamma distribution (to match the gamma process parametrization), for which the density is given by

$$\mathsf{Gam}(x; a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}.$$

Section 4.6 provides additional example applications of our theoretical results for two other CRMs: the beta process $\text{BP}(\gamma, \alpha, d)$ [Teh and Görür, 2009, Broderick et al., 2012] with Bernoulli or negative binomial likelihood, and the beta prime process $\text{BPP}(\gamma, \alpha, d)$ [Broderick et al., 2017] with odds-Bernoulli likelihood.

3.2 Sequential representations

Sequential representations are at the heart of the study of truncated CRMs. They provide an iterative method that can be terminated at any point to yield a finite approximation to the infinite process, where the choice of termination point determines the accuracy of the approximation. Thus, the natural first step in providing a coherent treatment of truncation analysis is to do the same for sequential representations. In past work, two major classes of sequential representation have been used: *series representations* of the form $\sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$, and *superposition representations* of the form $\sum_{k=1}^{\infty} \sum_{i=1}^{C_k} \theta_{ki} \delta_{\psi_{ki}}$, where each inner sum of C_k atoms is itself a CRM. This section examines four series representations [Ferguson and Klass, 1972, Bondesson, 1982, Rosiński, 1990, 2001] and three superposition representations (two of which are novel) [Broderick et al., 2012, 2017, James, 2014]. We show how previously-developed sequential representations for specific CRMs fit into these seven general representations. Finally, we discuss a stochastic mapping procedure that is useful in obtaining new representations from the transformation of others. Proofs for the results in this section may be found in Appendix A.2.

3.2.1 Series representations

Series representations arise from the transformation of a homogeneous Poisson point process [Rosiński, 2001]. They tend to be somewhat difficult to analyze due to the dependence between the atoms but also tend to produce very simple representations with small truncation error (cf. Section 4.3 and Chapter 5). Throughout the paper we let $\Gamma_k = \sum_{\ell=1}^{k} E_\ell$, $E_\ell \overset{\text{i.i.d.}}{\sim} \mathsf{Exp}(1)$, be the ordered jumps of a unit-rate homogeneous Poisson process on $\mathbb{R}_+$, let ν be a measure on $\mathbb{R}_+$ satisfying the basic conditions in Eq. (1), and let $\psi_k \overset{\text{i.i.d.}}{\sim} G$.

Inverse-Lévy [Ferguson and Klass, 1972] Define $\nu^{\leftarrow}(u) := \inf \{ x : \nu([x, \infty)) \le u \}$, the inverse of the tail measuer $\nu([x, \infty))$. We say Θ has an *inverse-Lévy* representation and write $\Theta \leftarrow \text{IL-Rep}(\nu)$ if

$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}, \qquad \text{with} \qquad \theta_k = \nu^{\leftarrow}(\Gamma_k).$$

Ferguson and Klass [1972] showed that $\Theta \leftarrow \text{IL-Rep}(\nu)$ implies $\Theta \sim \text{CRM}(\nu)$. The inverse-Lévy representation is analogous to the inverse CDF method for generating an arbitrary random variable from a uniform random variable, with the homogenous Poisson process playing the role of the uniform random variable. It is also the optimal sequential representation in the sense that the sequence θ_k that it generates is non-increasing. While an elegant and general approach, simulating the inverse-Lévy representation is difficult, as inverting the function $\nu([x, \infty))$ is analytically intractable except in a few cases.

Example 4.2.1 (Gamma process, $\Gamma P(\gamma, \lambda, 0)$)**.** We have $\nu([x, \infty)) = \gamma \lambda E_1(\lambda x)$, where $E_1(x) := \int_x^{\infty} u^{-1} e^{-u} \, du$ is the exponential integral function [Abramowitz and Stegun, 1964]. The inverse-Lévy representation for $\Gamma P(\gamma, \lambda, 0)$ is thus

$$\Theta = \sum_{k=1}^{\infty} \lambda^{-1} E_1^{-1}(\gamma^{-1} \lambda^{-1} \Gamma_k) \delta_{\psi_k}.$$

Neither E_1 nor its inverse can be computed in closed form, so one must resort to numerical approximations.

Bondesson [Bondesson, 1982]   We say Θ has a *Bondesson representation* and write $\Theta \leftarrow$ B-Rep(c, g) if for $c > 0$ and g a density on $\mathbb{R}_+$,

$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}, \qquad \text{with} \qquad \theta_k = V_k e^{-\Gamma_k/c}, \qquad V_k \overset{\text{i.i.d.}}{\sim} g.$$

Theorem 4.2.1 shows that Bondesson representations can be constructed for a large, albeit restricted, class of CRM rate measures. We offer a novel proof of Theorem 4.2.1 in Appendix A.2 using the induction strategy introduced by Banjevic et al. [2002]. Similar proof ideas are also used to prove truncation error bounds for sequential representations in Section 4.3. We use a slight abuse of notation for brevity: if $\nu(\mathrm{d}\theta)$ is a measure on $\mathbb{R}_+$ that is absolutely continuous with respect to Lebesgue measure, then $\nu(\theta)$ is the density of $\nu(\mathrm{d}\theta)$ with respect to the Lebesgue measure.

Theorem 4.2.1 (Bondesson representation [Bondesson, 1982])**.** *Let* $\nu(\mathrm{d}\theta) = \nu(\theta)\mathrm{d}\theta$ *be a rate measure satisfying Eq.* (1)*. If* $\theta\nu(\theta)$ *is nonincreasing,* $\lim_{\theta \to \infty} \theta\nu(\theta) = 0$*, and* $c_\nu := \lim_{\theta \to 0} \theta\nu(\theta) < \infty$*, then* $g_\nu(v) := -c_\nu^{-1} \frac{\mathrm{d}}{\mathrm{d}v}[v\nu(v)]$ *is a density on* $\mathbb{R}_+$ *and*

$$\Theta \leftarrow \text{B-Rep}(c_\nu, g_\nu) \quad \textit{implies} \quad \Theta \sim \text{CRM}(\nu).$$

Example 4.2.2 (Bondesson representation for $\Gamma\mathrm{P}(\gamma, \lambda, 0)$)**.** The following representation for the gamma process with $d = 0$ was described by Bondesson [1982] and Banjevic et al. [2002]. Since $\theta\nu(\theta) = \gamma\lambda e^{-\lambda\theta}$ is non-increasing and $c_\nu = \lim_{\theta \to 0} \theta\nu(\theta) = \gamma\lambda$, we obtain $g_\nu(v) = \lambda e^{-\lambda v} = \text{Exp}(v; \lambda)$. Thus, it follows from Theorem 4.2.1 that if $\Theta \leftarrow \text{B-Rep}(\gamma\lambda, \text{Exp}(\lambda))$, then $\Theta \sim \Gamma\mathrm{P}(\gamma, \lambda, 0)$. The condition that $\theta\nu(\theta)$ is non-increasing fails to hold if $d > 0$, so we cannot apply Theorem 4.2.1 to $\Gamma\mathrm{P}(\gamma, \lambda, d)$ when $d > 0$.

Thinning [Rosiński, 1990]   Using the nomenclature of Rosiński [2001], we say Θ has a *thinning representation* and write $\Theta \leftarrow \text{T-Rep}(\nu, g)$ if g is a probability measure on $\mathbb{R}_+$ such that ν is absolutely continuous with respect to g, i.e. $\nu \ll g$, and

$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}, \qquad \text{with} \qquad \theta_k = V_k \mathbb{1}\left(\frac{\mathrm{d}\nu}{\mathrm{d}g}(V_k) \geq \Gamma_k\right), \qquad V_k \overset{\text{i.i.d.}}{\sim} g.$$

Rosiński [1990] showed that $\Theta \leftarrow \text{T-Rep}(\nu, g)$ implies $\Theta \sim \text{CRM}(\nu)$. Note that $\Gamma_k \overset{a.s.}{\to} \infty$ as $k \to \infty$, so the probability that $\frac{\mathrm{d}\nu}{\mathrm{d}g}(V_k) \geq \Gamma_k$ is decreasing in k. Thus, this representation generates atoms with $\theta_k = 0$ (which have no effect and can be removed) increasingly frequently and becomes inefficient as $k \to \infty$.

Example 4.2.3 (Thinning representation for $\Gamma\mathrm{P}(\gamma, \lambda, d)$)**.** If we let $g = \text{Gam}(1 - d, \lambda)$, then the thinning representation for $\Gamma\mathrm{P}(\gamma, \lambda, d)$ is

$$\Theta = \sum_{k=1}^{\infty} V_k \mathbb{1}(V_k \Gamma_k \leq \gamma)\delta_{\psi_k}, \qquad \text{with} \qquad V_k \overset{\text{i.i.d.}}{\sim} \text{Gam}(1 - d, \lambda).$$

8

Rejection [Rosiński, 2001] Using the nomenclature of Rosiński [2001], we say Θ has a *rejection representation* and write $\Theta \leftarrow \text{R-Rep}(\nu, \mu)$ if μ is a measure on $\mathbb{R}_+$ satisfying Eq. (1) and $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \leq 1$, and

$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}, \qquad \text{with} \qquad \theta_k = V_k \mathbb{1}\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(V_k) \geq U_k\right), \qquad (V_k)_{k \in \mathbb{N}} \sim \text{PoissP}(\mu),$$

$$U_k \overset{\text{i.i.d.}}{\sim} \text{Unif}[0, 1].$$

Rosiński [2001] showed that $\Theta \leftarrow \text{R-Rep}(\nu, \mu)$ implies $\Theta \sim \text{CRM}(\nu)$. This representation is very similar to the thinning representation, except that the sequence $(V_k)_{k \in \mathbb{N}}$ is generated from a Poisson process on $\mathbb{R}_+$ rather than i.i.d. This allows $V_k \overset{a.s.}{\to} 0$ as $k \to \infty$, causing the frequency of generating ineffective atoms $\theta_k = 0$ to decay as $k \to \infty$, assuming μ is appropriately chosen such that $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\theta) \to 1$ as $\theta \to 0$. This representation can thus be constructed to be more efficient than the thinning representation. We can calculate the efficiency in terms of the expected number of rejections (that is, the number of θ_k that are identically zero):

Proposition 4.2.2. *For* $\text{R-Rep}(\nu, \mu)$*, the expected number of rejections is*

$$\mathbb{E}\left[\sum_{k=1}^{\infty} \mathbb{1}(\theta_k = 0)\right] = \int \left(1 - \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)\right) \mu(\mathrm{d}x).$$

Remark. If μ and ν can be written as densities with respect to Lebesgue measure, then the integral in Proposition 4.2.2 can be rewritten as $\int (\mu(x) - \nu(x))\mathrm{d}x$.

Example 4.2.4 (Rejection representation for $\Gamma\text{P}(\gamma, \lambda, 0)$)**.** Following Rosiński [2001], consider $\mu(\mathrm{d}\theta) = \gamma\lambda\theta^{-1}(1 + \lambda\theta)^{-1}\mathrm{d}\theta$. We call $\text{CRM}(\mu)$ the Lomax process, $\text{LomP}(\gamma, \lambda^{-1})$, after the related Lomax distribution. We can use the inverse-Lévy method analytically with μ since $\mu^{\leftarrow}(u) = \frac{1}{\lambda(e^{(\gamma\lambda)^{-1}u} - 1)}$. Thus, the rejection representation of $\Gamma\text{P}(\gamma, \lambda, 0)$ is

$$\Theta = \sum_{k=1}^{\infty} V_k \mathbb{1}(U_k \leq (1 + \lambda V_k)e^{-\lambda V_k})\delta_{\psi_k}, \quad \text{with} \quad V_k = \frac{1}{\lambda(e^{(\gamma\lambda)^{-1}\Gamma_k} - 1)}, \quad U_k \overset{\text{i.i.d.}}{\sim} \text{Unif}[0, 1].$$

Unlike in the thinning construction given in Example 4.2.3, only a finite number of rates will be set to zero almost surely. In particular, the expected number of rejections is $\gamma\lambda c_\gamma$, where c_γ is the Euler-Mascheroni constant.

Example 4.2.5 (Rejection representation for $\Gamma\text{P}(\gamma, \lambda, d)$, $d > 0$)**.** For the case of $d > 0$, we instead use $\mu(\mathrm{d}\theta) = \gamma\frac{\lambda^{1-d}}{\Gamma(1-d)}\theta^{-1-d}\mathrm{d}\theta$. We can again use the inverse-Lévy method analytically with μ since $\mu^{\leftarrow}(u) = (\gamma'u^{-1})^{1/d}$, where $\gamma' := \gamma\frac{\lambda^{1-d}}{d\Gamma(1-d)}$. The rejection representation is then

$$\Theta = \sum_{k=1}^{\infty} V_k \mathbb{1}(U_k \leq e^{-\lambda V_k})\delta_{\psi_k}, \qquad \text{with} \qquad V_k = (\gamma'\Gamma_k^{-1})^{1/d}, \qquad U_k \overset{\text{i.i.d.}}{\sim} \text{Unif}[0, 1].$$

The expected number of rejections is $\gamma\frac{\lambda^{1-d}}{d}$, so the representation is efficient for large d, but extremely inefficient when d is small.

3.2.2 Superposition representations

Superposition representations arise as an infinite sum of CRMs with finite rate measure. These tend to be easier to analyze than series representations as they decouple atoms between the summed CRMs, but can produce representations with larger truncation error (cf. Section 4.3 and Chapter 5). Throughout, let ν be a measure on $\mathbb{R}_+$ satisfying the basic conditions in Eq. (1), and let $\psi_k \overset{\text{i.i.d.}}{\sim} G$.

Decoupled Bondesson We say Θ has a *decoupled Bondesson representation* and write $\Theta \leftarrow$ DB-Rep(c, g, ξ) if for $c > 0, \xi > 0$, and g a density on $\mathbb{R}_+$,

$$\Theta = \sum_{k=1}^{\infty} \sum_{i=1}^{C_k} \theta_{ki} \delta_{\psi_{ki}}, \qquad \text{with} \qquad C_k \overset{\text{i.i.d.}}{\sim} \mathsf{Poiss}(c/\xi), \qquad \theta_{ki} = V_{ki} e^{-T_{ki}}, \qquad (3)$$

$$T_{ki} \overset{\text{indep}}{\sim} \mathsf{Gam}(k, \xi), \qquad V_{ki} \overset{\text{i.i.d.}}{\sim} g.$$

This is a novel superposition representation, though special cases are already known [Paisley et al., 2010, Roychowdhury and Kulis, 2015]. Theorem 4.2.3 shows that the decoupled Bondesson representation applies to the same class of CRMs as the Bondesson representation from Section 4.2.1.

Theorem 4.2.3 (Decoupled Bondesson representation). *Let* $\nu(\mathrm{d}\theta) = \nu(\theta)\mathrm{d}\theta$, c_ν, *and* g_ν *be as specified in Theorem 4.2.1. Then for any fixed* $\xi > 0$,

$$\Theta \leftarrow \mathrm{DB\text{-}Rep}(c_\nu, g_\nu, \xi) \quad \text{implies} \quad \Theta \sim \mathrm{CRM}(\nu).$$

The proof of Theorem 4.2.3 in Appendix A.2 generalizes the arguments from Paisley et al. [2010] and Roychowdhury and Kulis [2015]. The free parameter ξ controls the number of atoms generated for each outer sum index k; its principled selection can be made by trading off computational complexity (cf. Section 4.5) and truncation error (cf. Section 4.3).

Example 4.2.6 (Decoupled Bondesson representation for $\Gamma\mathrm{P}(\gamma, \lambda, 0)$). Arguments paralleling those made in Example 4.2.2 show that the $\Gamma\mathrm{P}(\gamma, \lambda, 0)$ representation from Roychowdhury and Kulis [2015] follows directly from an application of Theorem 4.2.3: if $\Theta \leftarrow$ DB-Rep$(\gamma\lambda, \mathsf{Exp}(\lambda), \xi)$, then $\Theta \sim \Gamma\mathrm{P}(\gamma, \lambda, 0)$. As in the Bondesson representation setting, Theorem 4.2.3 does not apply to $\Gamma\mathrm{P}(\gamma, \lambda, d)$ when $d > 0$ because the condition that $\theta\nu(\theta)$ is non-increasing fails to hold.

Size-biased [Broderick et al., 2017, James, 2014] Let $\pi(\theta) := h(0 \,|\, \theta)$. We say Θ has a *size-biased representation* and write $\Theta \leftarrow$ SB-Rep(ν, h) if

$$\Theta = \sum_{k=1}^{\infty} \sum_{i=1}^{C_k} \theta_{ki} \delta_{\psi_{ki}}, \qquad \text{with} \qquad C_k \overset{\text{indep}}{\sim} \mathsf{Poiss}\left(\eta_k\right),$$

$$\theta_{ki} \overset{\text{indep}}{\sim} \frac{1}{\eta_k} \pi(\theta)^{k-1} \left(1 - \pi(\theta)\right) \nu(\mathrm{d}\theta), \qquad (4)$$

$$\eta_k := \int \pi(\theta)^{k-1} \left(1 - \pi(\theta)\right) \nu(\mathrm{d}\theta).$$

Broderick et al. [2017] and James [2014] showed that $\Theta \leftarrow$ SB-Rep(ν, h) implies $\Theta \sim \mathrm{CRM}(\nu)$. If the rate measure ν and the likelihood h are selected to be a conjugate exponential family then,

noting that $\sum_{x=1}^{\infty} h(x \mid \theta) = 1 - \pi(\theta)$, the rate θ_{ki} can be sampled from a mixture of exponential family distributions:

$$\theta_{ki} \mid z_{ki} \overset{\text{indep}}{\sim} \frac{1}{\eta_{kz_{ki}}} h(z_{ki} \mid \theta) \pi(\theta)^{k-1} \nu(\mathrm{d}\theta), \qquad z_{ki} \overset{\text{indep}}{\sim} \mathsf{Categorical}\left((\eta_{kx}/\eta_k)_{x=1}^{\infty}\right),$$

$$\eta_{kx} := \int h(x \mid \theta) \pi(\theta)^{k-1} \nu(\mathrm{d}\theta).$$

Example 4.2.7 (Size-biased representation for $\Gamma\mathrm{P}(\gamma, \lambda, d)$)**.** For the Gamma process, values for η_{kx} and η_k can be found using integration by parts and the standard gamma distribution integral, while $\theta_{ki} \mid z_{ki}$ is sampled from a gamma distribution by inspection:

$$\eta_{kx} = \frac{\gamma \lambda^{1-d} \Gamma(x-d)}{x! \Gamma(1-d)(\lambda+k)^{x-d}}, \qquad \eta_k = \begin{cases} \frac{\gamma \lambda^{1-d}}{d}\left((\lambda+k)^d - (\lambda+k-1)^d\right) & d > 0 \\ \gamma\lambda\left(\log(\lambda+k) - \log(\lambda+k-1)\right) & d = 0 \end{cases},$$

$$\theta_{ki} \mid z_{ki} \overset{\text{indep}}{\sim} \mathsf{Gam}(x-d, \lambda+k).$$

Power-law   We say Θ has a *power-law representation* and write $\Theta \leftarrow \mathrm{PL\text{-}Rep}(\gamma, \alpha, d, g)$ if for $\gamma > 0, 0 \leq d < 1, \alpha > -d$, and g a density on $\mathbb{R}_+$,

$$\Theta = \sum_{k=1}^{\infty} \sum_{i=1}^{C_k} \theta_{ki} \delta_{\psi_{ki}}, \qquad \text{with} \qquad C_k \overset{\text{i.i.d.}}{\sim} \mathsf{Poiss}(\gamma), \qquad \theta_{ki} = V_{ki} U_{kik} \prod_{j=1}^{k-1}(1 - U_{kij}), \qquad (5)$$

$$V_{ki} \overset{\text{i.i.d.}}{\sim} g, \qquad\qquad U_{kij} \overset{\text{indep}}{\sim} \mathsf{Beta}(1-d, \alpha+jd).$$

This is a novel superposition representation, although it was previously developed in the special case of the beta process (where $g(v) = \delta_1$) [Broderick et al., 2012]. The name of this representation arises from the fact that it exhibits Types I and II power-law behavior [Broderick et al., 2012] under mild conditions when $d > 0$, as we show in Theorem A.2.1 in the appendix (note, however, that it will not exhibit power-law behavior when $d = 0$). Theorem 4.2.5 below shows the conditions under which $\Theta \leftarrow \mathrm{PL\text{-}Rep}(\gamma, \alpha, d, g)$ implies $\Theta \sim \mathrm{CRM}(\nu)$. Its proof in Appendix A.2 relies on the notion of *stochastic mapping* (Lemma 4.2.4), a powerful technique for transforming one CRM into another. Note that in Lemma 4.2.4, the case where u is a deterministic function of θ via the mapping $u = \tau(\theta)$ may be recovered by setting $\kappa(\theta, \mathrm{d}u) = \delta_{\tau(\theta)}$.

Lemma 4.2.4 (CRM stochastic mapping)**.** *Let $\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k} \sim \mathrm{CRM}(\nu)$. Then for any probability kernel $\kappa(\theta, \mathrm{d}u)$, we have $\kappa(\Theta) \sim \mathrm{CRM}(\nu_\kappa)$, where*

$$\kappa(\Theta) := \sum_{k=1}^{\infty} u_k \delta_{\psi_k}, \qquad u_k \mid \theta_k \sim \kappa(\theta_k, \cdot), \qquad \text{and} \qquad \nu_\kappa(\mathrm{d}u) := \int \kappa(\theta, \mathrm{d}u) \nu(\mathrm{d}\theta).$$

Theorem 4.2.5 (Power-law representation)**.** *Let $\nu(\mathrm{d}\theta) = \nu(\theta)\mathrm{d}\theta$ be a rate measure satisfying Eq. (1), and let g_ν be a density on $\mathbb{R}_+$ such that*

$$\nu(u) = \int \theta^{-1} g_\nu \left(u\theta^{-1}\right) \nu_{\mathrm{BP}}(\mathrm{d}\theta),$$

where

$$\nu_{\mathrm{BP}}(\mathrm{d}\theta) = \gamma \frac{\Gamma(\alpha+1)}{\Gamma(1-d)\Gamma(\alpha+d)} \mathbb{1}\left[\theta \le 1\right] \theta^{-1-d}(1-\theta)^{\alpha+d-1}\mathrm{d}\theta$$

is the rate measure for the beta process $\mathrm{BP}(\gamma,\alpha,d)$ *from Eq.* (24)*. Then*

$$\Theta \leftarrow \mathrm{PL\text{-}Rep}(\gamma,\alpha,d,g_\nu) \quad implies \quad \Theta \sim \mathrm{CRM}(\nu).$$

Example 4.2.8 (Power-law representation for $\Gamma\mathrm{P}(\gamma,\lambda,d)$)**.** If we choose $g_\nu = \mathsf{Gam}(\lambda,\lambda)$, then using the change of variable $w = u(\theta^{-1}-1)$,

$$\int \theta^{-1}g_\nu\left(u\theta^{-1}\right)\nu_{\mathrm{BP}}\left(\mathrm{d}\theta\right)$$

$$= \gamma\lambda\frac{\lambda^\lambda}{\Gamma(1-d)\Gamma(\lambda+d)}u^{\lambda-1}\int \theta^{-\lambda-d-1}e^{-\lambda u\theta^{-1}}(1-\theta)^{\lambda+d-1}\,\mathrm{d}\theta\,\mathrm{d}u$$

$$= \gamma\lambda\frac{\lambda^\lambda}{\Gamma(1-d)\Gamma(\lambda+d)}u^{-1-d}e^{-\lambda u}\int w^{\lambda+d-1}e^{-\lambda w}\,\mathrm{d}w\,\mathrm{d}u$$

$$= \gamma\frac{\lambda^{1-d}}{\Gamma(1-d)}u^{-1-d}e^{-\lambda u}\,\mathrm{d}u.$$

It follows immediately from Theorem 4.2.5 that if $\Theta \leftarrow \mathrm{PL\text{-}Rep}(\gamma,\lambda,d,\mathsf{Gam}(\lambda,\lambda))$, then $\Theta \sim \Gamma\mathrm{P}(\gamma,\lambda,d)$. To the best knowledge of the authors, this power-law representation for the gamma process is novel.

3.3 Truncation analysis

Each of the sequential representations developed in Section 4.2 shares a common structural element—an outer infinite sum—which is responsible for generating a countably infinite number of atoms in the CRM. In this section, we terminate these outer sums at a finite *truncation level* $K \in \mathbb{N}$, resulting in a *truncated CRM* Θ_K possessing a finite number of atoms. We develop upper bounds on the error induced by this truncation procedure. All of the truncated CRM error bounds in this section rely on Lemma 4.3.1, which is a tightening (by a factor of two) of the bound in Ishwaran and James [2001, 2002] (for its generalization to arbitrary discrete random measures, see Lemma A.3.1). Proposition 4.3.2 shows that the bound in Lemma 4.3.1 is tight without further assumptions on the data likelihood f.

Lemma 4.3.1 (CRM protobound)**.** *Let* $\Theta \sim \mathrm{CRM}(\nu)$. *For any truncation* Θ_K, *if*

$$X_n \,|\, \Theta \overset{i.i.d.}{\sim} \mathrm{LP}(h,\Theta), \qquad\qquad Z_n \,|\, \Theta_K \overset{i.i.d.}{\sim} \mathrm{LP}(h,\Theta_K),$$

$$Y_n \,|\, X_n \overset{indep}{\sim} f(\cdot \,|\, X_n), \qquad\qquad W_n \,|\, Z_n \overset{indep}{\sim} f(\cdot \,|\, Z_n),$$

then, with $p_{N,\infty}$ *and* $p_{N,K}$ *denoting the marginal densities of* $Y_{1:N}$ *and* $W_{1:N}$, *respectively,*

$$\frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 \le 1 - \mathbb{P}\left(\mathrm{supp}(X_{1:N}) \subseteq \mathrm{supp}(\Theta_K)\right),$$

Proposition 4.3.2 (Protobound tightness). *If G is non-atomic and Ψ is Borel, then for any $\delta > 0$, there exists a likelihood f such that Lemma 4.3.1 is tight up to a factor of $1 - \delta$.*

The proof of all results in this section (including Lemma 4.3.1 and Proposition 4.3.2) can be found in Appendix A.3. All of the provided truncation results use the generative model in Lemma 4.3.1, and are summarized in Table 1 in Chapter 5. Throughout this section, for a given likelihood model $h(x \,|\, \theta)$ we define $\pi(\theta) := h(0 \,|\, \theta)$ for notational brevity. The asymptotic behavior of truncation error bounds is specified with tilde notation:

$$a(K) \sim b(K), \quad K \to \infty \quad \Longleftrightarrow \quad \lim_{K \to \infty} \frac{a(K)}{b(K)} = 1.$$

3.3.1 Series representations

Each of the series representations can be viewed a functional of a standard Poisson point process and a sequence of i.i.d. random variables with some distribution g on $\mathbb{R}_+$. In particular, we may write each in the form

$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}, \qquad \text{with} \qquad \theta_k = \tau(V_k, \Gamma_k), \qquad V_k \overset{\text{i.i.d.}}{\sim} g, \qquad (6)$$

where Γ_k are the jumps of a unit-rate homogeneous Poisson point process on $\mathbb{R}_+$, and $\tau : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is a non-negative measurable function such that $\lim_{u \to \infty} \tau(v, u) = 0$ for g-almost every v. The truncated CRM then takes the form

$$\Theta_K := \sum_{k=1}^{K} \theta_k \delta_{\psi_k}.$$

Theorem 4.3.3 provides a general truncation error bound for series representations of the form Eq. (6), specifies its range, and guarantees that the bound decays to 0 as $K \to \infty$.

Theorem 4.3.3 (Series representation truncation error). *The error in approximating a series representation of Θ with its truncation Θ_K satisfies*

$$0 \le \frac{1}{2} \| p_{N,\infty} - p_{N,K} \|_1 \le 1 - e^{-B_{N,K}} \le 1,$$

where

$$B_{N,K} := \int_0^{\infty} \left(1 - \mathbb{E} \left[\pi \left(\tau \left(V, u + G_K \right) \right)^N \right] \right) \, \mathrm{d}u, \qquad (7)$$

$G_0 := 0$, $G_K \overset{\text{indep}}{\sim} \mathsf{Gam}(K, 1)$ *for $K \ge 1$, and $V \overset{\text{indep}}{\sim} g$. Furthermore, $\forall N \in \mathbb{N}$, $\lim_{K \to \infty} B_{N,K} = 0$.*

Remark. An alternate form of $B_{N,K}$ that is sometimes easier to use in practice can be found by applying the standard geometric series formula to Eq. (7), which yields

$$B_{N,K} = \sum_{n=1}^{N} \int_0^{\infty} \mathbb{E} \left[\pi \left(\tau \left(V, u + G_K \right) \right)^{n-1} \left(1 - \pi \left(\tau \left(V, u + G_K \right) \right) \right) \right] \, \mathrm{d}u.$$

A simplified upper bound on $B_{N,K}$ can be derived by noting that $\pi(\theta) \leq 1$, so

$$B_{N,K} \leq N \int_0^\infty \left(1 - \mathbb{E}\left[\pi\left(\tau\left(V, u + G_K\right)\right)\right]\right) \, \mathrm{d}u. \tag{8}$$

This bound usually gives the same asymptotics in K as Eq. (7).

The main task in using Theorem 4.3.3 to develop a truncation error bound for a series representation is evaluating the integrand in the definition of $B_{N,K}$. Thus, we next evaluate the integrand and provide expressions of the truncation error bound for the four series representations outlined in Section 4.2.1. Throughout the remainder of this section, G_K is defined as in Theorem 4.3.3, $F_0 \equiv 1$, and F_K is the CDF of G_K.

Inverse-Lévy representation For this representation we have

$$\tau(v, u) = \nu^\leftarrow(u) := \inf\left\{y : \nu\left([y, \infty)\right) \leq u\right\}.$$

To evaluate the bound in Eq. (8), we use the transformation of variables $x = \nu^\leftarrow(u + G_K)$ and the fact that for $a, b \geq 0$, $\nu^\leftarrow(a) \geq b \iff a \leq \nu\left([b, \infty)\right)$ to conclude that

$$B_{N,K} \leq N \int_0^\infty F_K(\nu[x, \infty))(1 - \pi(x)) \, \nu(\mathrm{d}x). \tag{9}$$

Recent work on the inverse-Lévy representation has developed Monte Carlo estimates of the error of the truncated random measure moments for those $\nu\left([x, \infty)\right)$ with known inverse $\nu^\leftarrow$ [Arbel and Prünster, 2017]. In contrast, the result above provides an explicit bound on the L^1 truncation error. Our bound does not require knowing $\nu^\leftarrow$, which is often the most challenging aspect of applying the inverse-Lévy representation.

Example 4.3.1 (IL-Rep truncation for $\mathrm{LomP}(\gamma, \lambda^{-1})$ with Poisson likelihood)**.** Recall from Example 4.2.4 that the Lomax process $\mathrm{LomP}(\gamma, \lambda^{-1})$ is the CRM with rate measure $\nu(\mathrm{d}\theta) = \gamma\lambda\theta^{-1}(1 + \lambda\theta)^{-1}\mathrm{d}\theta$, so $\nu[x, \infty) = \gamma\lambda\log\{1 + (\lambda x)^{-1}\}$. Using Eq. (9), we have

$$B_{N,K} \leq N\gamma\lambda \int_0^\infty F_K(\gamma\lambda\log\{1 + (\lambda x)^{-1}\})(1 - e^{-x})x^{-1}(1 + \lambda x)^{-1}\mathrm{d}x.$$

Since $F_K(t) \leq t^K/K! \leq (3t/K)^K$, for any $a > 0$ the integral is upper bounded by

$$\int_0^a (1 + \lambda x)^{-1}\mathrm{d}x + F_K\left(\gamma\lambda\log\left\{1 + \frac{1}{\lambda a}\right\}\right) \int_a^\infty x^{-1}(1 + \lambda x)^{-1}\mathrm{d}x$$

$$\leq a + F_K\left(\gamma\lambda\log\left\{1 + \frac{1}{\lambda a}\right\}\right)\log\left\{1 + \frac{1}{\lambda a}\right\}$$

$$\leq \lambda^{-1}(e^b - 1)^{-1} + b(3\gamma\lambda b/K)^K \quad \text{where} \quad b := \log\{1 + (\lambda a)^{-1}\}. \tag{10}$$

Replacing $(e^b - 1)^{-1}$ with the approximation e^{-b} and then setting the two terms in Eq. (10) equal, we obtain $b = KW_0\left(\left\{3\gamma\lambda^{\frac{K+2}{K+1}}(K+1)^{\frac{1}{K+1}}\right\}^{-1}\right)$, where W_0 is the product logarithm function, i.e.

$$W_0(y) = x \iff xe^x = y. \tag{11}$$

Thus, using the fact that $e^{-t} \leq (e^t - 1)^{-1}$ and $\lambda^{\frac{K+2}{K+1}}(K+1)^{\frac{1}{K+1}}$ reaches its maximum at $K = \max(0, e\lambda^{-1} - 1)$, we conclude that

$$B_{N,K} \leq \frac{2N\gamma[1 + (3\gamma\lambda)^{-1}]}{\exp\left(KW_0\left(\{3\gamma\lambda\max(\lambda, e)\}^{-1}\right)\right) - 1}$$
$$\sim 2N\gamma[1 + (3\gamma\lambda)^{-1}]e^{-KW_0(\{3\gamma\lambda\max(\lambda,e)\}^{-1})} \qquad K \to \infty.$$

Bondesson representation   For this representation we have

$$\tau(v, u) = ve^{-u/c}, \qquad\qquad g(\mathrm{d}v) = -c^{-1}\frac{\mathrm{d}}{\mathrm{d}v}(v\nu(v))\,\mathrm{d}v.$$

Writing the expectation over V explicitly as an integral with measure $g(v)\mathrm{d}v$, using the transformation of variables $u = -c\log x/v$ (so $x = ve^{-u/c}$), and given the definition of $g(v) = -c^{-1}\frac{\mathrm{d}}{\mathrm{d}v}(v\nu(v))$ for the Bondesson representation, we have

$$B_{N,K} \leq N\int_0^\infty \left(1 - \mathbb{E}\left[\pi\left(ve^{-G_K/c}\right)\right]\right)\nu(\mathrm{d}v).$$

Example 4.3.2 (Truncation of the Bondesson representation for $\Gamma\mathrm{P}(\gamma, \lambda, 0)$)**.** Let $\tilde{G}_K \overset{\mathcal{D}}{=} G_K/(\gamma\lambda)$. Since $\pi(\theta) = e^{-\theta}$ and $c = \gamma\lambda$, we have

$$\int_0^\infty \left(1 - \mathbb{E}\left[\pi(ve^{-G_K/c})\right]\right)\nu(\mathrm{d}v) = \gamma\lambda\,\mathbb{E}\left[\int_0^\infty (1 - e^{-ve^{-\tilde{G}_K}})v^{-1}e^{-\lambda v}\mathrm{d}v\right]$$
$$= \gamma\lambda\,\mathbb{E}\left[\log(1 + e^{-\tilde{G}_K}/\lambda)\right]$$
$$\leq \gamma\,\mathbb{E}\left[e^{-\tilde{G}_K}\right] = \gamma\left(\frac{\gamma\lambda}{1 + \gamma\lambda}\right)^K.$$

The second equality follows by using the power series for the exponential integral [Abramowitz and Stegun, 1964, Chapter 5]. Thus,

$$B_{N,K} \leq N\gamma\left(\frac{\gamma\lambda}{1 + \gamma\lambda}\right)^K.$$

Thinning representation   For this representation we have

$$\tau(v, u) = v\mathbb{1}\left[\frac{\mathrm{d}\nu}{\mathrm{d}g}(v) \geq u\right], \qquad\qquad g \text{ any distribution on } \mathbb{R}_+ \text{ s.t. } \nu \ll g.$$

Since $\pi(0) = 1$ by Lemma A.1.3, we have that $1 - \pi(v\mathbb{1}(A)) = (1 - \pi(v))\mathbb{1}(A)$ for any event A. Using this fact, we have

$$B_{N,K} \leq N\int_0^\infty (1 - \pi(v))\int_0^{\frac{\mathrm{d}\nu}{\mathrm{d}g}(v)} F_K(u)\,\mathrm{d}u\,g(\mathrm{d}v). \tag{12}$$

Analytic bounds for the thinning representation of specific processes tend to be opaque and notationally cumbersome, so we simply compare its truncation error in Chapter 5 to the other representations by numerical approximation of Eq. (12).

Rejection representation Assume that we can use the inverse-Lévy representation to simulate $\mathrm{PoissP}(\mu)$. Then for the rejection representation we have

$$\tau(v,u) = \mu^{\leftarrow}(u)\mathbb{1}\left[\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\mu^{\leftarrow}(u)) \geq v\right], \qquad g(\mathrm{d}v) = \mathbb{1}[0 \leq v \leq 1]\mathrm{d}v,$$

where μ satisfies $\nu \ll \mu$, $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \leq 1$, and $\mu^{\leftarrow}(u) := \inf\{x : \mu([x,\infty)) \leq u\}$. Using the same techniques as for the thinning and inverse-Lévy representations, we have that

$$B_{N,K} \leq N \int_0^{\infty} F_K(\mu[x,\infty))(1 - \pi(x))\,\nu(\mathrm{d}x). \tag{13}$$

Example 4.3.3 (R-Rep truncation for $\Gamma\mathrm{P}(\gamma,\lambda,0)$ with Poisson likelihood)**.** Using Eq. (13) and the fact that $1 - e^{-x} \leq x$, we have

$$B_{N,K} \leq N\gamma\lambda \int_0^{\infty} F_K(\gamma\lambda \log\{1 + (\lambda x)^{-1}\})e^{-\lambda x}\mathrm{d}x. \tag{14}$$

Arguing as in Example 4.3.1, we see that the integral in Eq. (14) is upper bounded by

$$\int_0^a e^{-\lambda x}\mathrm{d}x + F_K\left(\gamma\lambda \log\left\{1 + \frac{1}{\lambda a}\right\}\right)\int_a^{\infty} e^{-\lambda x}\mathrm{d}x$$

$$\leq a + \lambda^{-1}F_K\left(\gamma\lambda \log\left\{1 + \frac{1}{\lambda a}\right\}\right)$$

$$= \lambda^{-1}\left((e^b - 1)^{-1} + (3\gamma\lambda b/K)^K\right), \tag{15}$$

where $b := \log\{1 + (\lambda a)^{-1}\}$. Replacing $(e^b - 1)^{-1}$ with the approximation e^{-b} and then setting the two terms in Eq. (15) equal to each other, we obtain $b = KW_0(\{3\gamma\lambda\}^{-1})$ (where W_0 is defined in Eq. (11)) and conclude that

$$B_{N,K} \leq \frac{2N\gamma}{e^{KW_0(\{3\gamma\lambda\}^{-1})} - 1} \sim 2N\gamma e^{-KW_0(\{3\gamma\lambda\}^{-1})} \qquad K \to \infty.$$

Example 4.3.4 (R-Rep truncation for $\Gamma\mathrm{P}(\gamma,\lambda,d)$ with Poisson likelihood, $d > 0$)**.** We have

$$B_{N,K} \leq N\frac{\gamma\lambda^{1-d}}{\Gamma(1-d)} \int_0^{\infty} F_K(\gamma'x^{-d})(1 - e^{-x})x^{-1-d}e^{-\lambda x}\mathrm{d}x.$$

The integral can be upper bounded as

$$\int_0^a x^{-d}\mathrm{d}x + F_K(\gamma'a^{-d})\int_a^{\infty}(1 + e^{-x})x^{-1-d}e^{-\lambda x}\mathrm{d}x$$

$$\leq (1-d)^{-1}a^{1-d} + \Gamma(-d)(\lambda^d - (1+\lambda)^d)(3\gamma'K^{-1}a^{-d})^K.$$

Setting the two terms equal and solving for a, we obtain

$$B_{N,K} \leq 2N\frac{\gamma\lambda^{1-d}}{\Gamma(2-d)}[(1-d)\Gamma(-d)]^{\frac{d(1-d)}{d(1-d)+K}}\left[\frac{3\gamma\lambda^{1-d}}{d\Gamma(1-d)K}\right]^{\frac{Kd(1-d)}{d(1-d)+K}}$$

$$\sim 2N\frac{\gamma\lambda^{1-d}}{\Gamma(2-d)}\left[\frac{3\gamma\lambda^{1-d}}{d\Gamma(1-d)}\right]^{d(1-d)}K^{-d(1-d)} \qquad K \to \infty.$$

16

3.3.2 Superposition representations

For superposition representations, the truncated CRM takes the form

$$\Theta_K := \sum_{k=1}^{K} \sum_{i=1}^{C_k} \theta_{ki} \delta_{\psi_{ki}}.$$

Let $\Theta_K^+ := \Theta - \Theta_K$ denote the *tail measure*. By the superposition property of Poisson point processes [Kingman, 1993], the tail measure is itself a CRM with some rate measure ν_K^+ and is independent of Θ_K:

$$\Theta_K^+ = \sum_{k=K+1}^{\infty} \sum_{i=1}^{C_k} \theta_{ki} \delta_{\psi_{ki}} \sim \text{CRM}\left(\nu_K^+\right), \qquad \Theta_K^+ \perp\!\!\!\perp \Theta_K, \qquad \Theta = \Theta_K + \Theta_K^+. \qquad (16)$$

The following result provides a general truncation error bound for superposition representations, specifies its range, and guarantees that the bound decays to 0 as $K \to \infty$.

Theorem 4.3.4 (Superposition representation truncation error)**.** *The error in approximating a superposition representation of $\Theta \sim \text{CRM}(\nu)$ with its truncation Θ_K satisfies*

$$0 \le \frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 \le 1 - e^{-B_{N,K}} \le 1,$$

where

$$B_{N,K} := \int \left(1 - \pi(\theta)^N\right) \nu_K^+(\mathrm{d}\theta). \qquad (17)$$

Furthermore, $\forall N \in \mathbb{N}$, $\lim_{K \to \infty} B_{N,K} = 0$.

Remark. As for series representations, an alternate form of $B_{N,K}$ that is sometimes easier to use can be found by applying the standard geometric series formula to Eq. (17):

$$B_{N,K} = \sum_{n=1}^{N} \int \pi(\theta)^{n-1} \left(1 - \pi(\theta)\right) \nu_K^+(\mathrm{d}\theta).$$

A simplified upper bound on $B_{N,K}$ can be derived by noting that $\pi(\theta) \le 1$, so

$$B_{N,K} \le N \int_0^\infty \left(1 - \pi\left(\theta\right)\right) \nu_K^+(\mathrm{d}\theta).$$

This bound usually gives the same asymptotics in K as Eq. (17).

The main task in using Theorem 4.3.4 to develop a truncation error bound for a superposition representation is determining its tail measure ν_K^+. In the following, we provide the tail measure for the three superposition representations outlined in Section 4.2.2.

Decoupled Bondesson representation For each point process in the superposition, an average of c/ξ atoms are generated with independent weights of the form $V e^{-T_k}$ where $V \overset{\text{indep}}{\sim} g$ and $T_k \overset{\text{indep}}{\sim}$ $\mathsf{Gam}(k, \xi)$. Therefore, the tail measure is

$$\nu_K^+(\mathrm{d}\theta) = \frac{c}{\xi} \sum_{k=K+1}^{\infty} \tilde{g}_{k,\xi}(\theta)\mathrm{d}\theta,$$

where $\tilde{g}_{k,\xi}$ is the density of $V e^{-T_k}$. The bound for the decoupled Bondesson representation can therefore be expressed as

$$B_{N,K} \leq N\frac{c}{\xi} \sum_{k=K+1}^{\infty} \mathbb{E}\left[1 - \pi\left(V e^{-T_k}\right)\right].$$

Example 4.3.5 (Decoupled Bondesson representation truncation for $\Gamma\mathrm{P}(\gamma, \lambda, 0)$). Using the fact that $1 - e^{-\theta} \leq \theta$, we have

$$B_{N,K} = \frac{N\gamma\lambda}{\xi} \sum_{k=K+1}^{\infty} \mathbb{E}[1 - \pi(V_{k1}e^{-T_{k1}})] \leq \frac{N\gamma\lambda}{\xi} \sum_{k=K+1}^{\infty} \mathbb{E}[V_{k1}e^{-T_{k1}}]$$

$$= \frac{N\gamma\lambda}{\xi} \sum_{k=K+1}^{\infty} \frac{1}{\lambda}\left(\frac{\xi}{1+\xi}\right)^k = N\gamma\left(\frac{\xi}{1+\xi}\right)^K,$$

which is equivalent (up to a factor of 2) to the bound in Roychowdhury and Kulis [2015].

Size-biased representation The constructive derivation of the size-biased representation [Broderick et al., 2017, proof of Theorem 5.1] immediately yields

$$\nu_K^+(\mathrm{d}\theta) = \pi(\theta)^K \nu(\mathrm{d}\theta).$$

Therefore, the size-biased representation truncation error bound can be expressed using the formula for η_k from Eq. (4) as

$$B_{N,K} = \sum_{n=1}^{N} \int \pi(\theta)^{K+n-1}(1 - \pi(\theta))\nu(\mathrm{d}\theta) = \sum_{n=1}^{N} \eta_{K+n}. \tag{18}$$

Example 4.3.6 (Size-biased representation truncation for $\Gamma\mathrm{P}(\gamma, \lambda, d)$). For $d > 0$, the standard gamma integral yields

$$\eta_k = \int \pi(\theta)^{k-1}(1 - \pi(\theta))\nu(\mathrm{d}\theta) = \frac{\gamma\lambda^{1-d}}{d}\left((\lambda + k)^d - (\lambda + k - 1)^d\right).$$

The sum from Eq. (18) is telescoping, so canceling terms,

$$B_{N,K} \leq \frac{\gamma\lambda^{1-d}}{d}\left((\lambda + K + N)^d - (\lambda + K)^d\right) \sim \gamma N\lambda^{1-d}K^{d-1} \qquad K \to \infty,$$

where the asymptotic result follows from Lemma A.1.4. To analyze the $d = 0$ case, we use L'Hospital's rule to take the limit of the integral:

$$\lim_{d \to 0} \int \pi(\theta)^{k-1}(1 - \pi(\theta))\nu(\mathrm{d}\theta) = \gamma\lambda\left(\log(\lambda + k) - \log(\lambda + k - 1)\right).$$

Canceling terms in the telescopic sum yields

$$B_{N,K} \leq \gamma\lambda\left(\log(\lambda + K + N) - \log(\lambda + K)\right) \sim \gamma\lambda N K^{-1} \qquad K \to \infty,$$

where the asymptotic result follows from an application of Lemma A.1.4.

Power-law representation For each point process in the superposition, an average of γ atoms are generated with independent weights of the form $V U_k \prod_{\ell=1}^{k-1}(1 - U_\ell)$, where $V \overset{\text{indep}}{\sim} g$ and $U_\ell \overset{\text{indep}}{\sim}$ Beta$(1 - d, \alpha + \ell d)$. Therefore, the tail measure is

$$\nu_K^+(\mathrm{d}\theta) = \gamma \sum_{k=K+1}^{\infty} \tilde{g}_k(\theta)\mathrm{d}\theta,$$

where $\tilde{g}_k$ is the density of the random variable $V U_k \prod_{\ell=1}^{k-1}(1 - U_\ell)$. The truncation error bound may be expressed as

$$B_{N,K} \leq N\gamma \sum_{k=K+1}^{\infty} \mathbb{E}\left[1 - \pi\left(V U_k \prod_{\ell=1}^{k-1}(1 - U_\ell)\right)\right].$$

Example 4.3.7 (Power-law representation truncation for $\Gamma\mathrm{P}(\gamma, \lambda, d)$). Let β_k be a random variable with density $\tilde{g}_k$ (with λ in the place of α). Using $1 - e^{-\theta} \leq \theta$, we have

$$\sum_{k=K+1}^{\infty} \mathbb{E}[1 - \pi(\beta_k)] \leq \sum_{k=K+1}^{\infty} \mathbb{E}[\beta_k] = \mathbb{E}\left[\sum_{k=K+1}^{\infty} \beta_k\right] = \prod_{k=1}^{K} \frac{\lambda + kd}{\lambda + kd - d + 1},$$

where the final equality follows from Ishwaran and James [2001, Theorem 1]. Thus,

$$B_{N,K} \leq \gamma N \prod_{k=1}^{K} \frac{\lambda + kd}{\lambda + kd - d + 1} \sim \gamma N \begin{cases} \left(\frac{\lambda}{\lambda+1}\right)^K & d = 0 \\ \frac{\Gamma\left(\frac{\lambda+1}{d}\right)}{\Gamma\left(\frac{\lambda+d}{d}\right)} K^{1-d^{-1}} & 0 < d < 1 \end{cases} \qquad K \to \infty, \qquad (19)$$

where the $0 < d < 1$ case in Eq. (19) follows by Lemma A.1.5 applied to

$$\prod_{k=1}^{K} \frac{\lambda + kd}{\lambda + kd - d + 1} = \frac{\Gamma((\lambda + 1)/d)}{\Gamma((\lambda + d)/d)} \frac{\Gamma(\lambda/d + K + 1)}{\Gamma(\lambda/d + K + d^{-1})}.$$

3.3.3 Stochastic mapping

We now show how truncation bounds developed elsewhere in this paper can be applied to CRM representations that have been transformed using Lemma 4.2.4. For $\Theta \sim \mathrm{CRM}(\nu)$, we denote its transformation by $\tilde{\Theta} = \kappa(\Theta)$. For any object defined with respect to Θ, the corresponding object for $\tilde{\Theta}$ is denoted with a tilde. For example, in place of N and $X_{1:N}$ (for Θ), we use $\tilde{N}$ and $\tilde{X}^{1:}_{\tilde{N}}$ (for $\tilde{\Theta}$). We make $B_{N,K}$ a function of $\pi(\theta)$ in the notation of Proposition 4.3.5; when one applies stochastic mapping to a CRM, one usually also wants to change the likelihood $h(x \,|\, \theta)$, and thus also changes $\pi(\theta) = h(0 \,|\, \theta)$. The proof of Proposition 4.3.5 may be found in Appendix A.3.

Proposition 4.3.5 (Truncation error under a stochastic mapping)**.** *Consider a representation for* $\Theta \sim \mathrm{CRM}(\nu)$ *with truncation error bound* $B_{N,K}(\pi)$*. Then for any likelihood* $\tilde{h}(x \,|\, u)$*, if* $\tilde{\Theta}$ *is a stochastic mapping of* Θ *under the probability kernel* $\kappa(\theta, \mathrm{d}u)$*, its truncation error bound is* $B_{1,K}(\pi_{\kappa,\tilde{N}})$*, where* $\pi_{\kappa,\tilde{N}}(\theta) := \int \tilde{h}(0 \,|\, u)^N \kappa(\theta, \mathrm{d}u)$*.*

3.3.4 Hyperpriors

In practice, prior distributions are often placed on the hyperparameters of the CRM rate measure (i.e. γ, α, λ, d, etc.). We conclude our investigation of CRM truncation error by showing how bounds developed in this section can be modified to account for the use of h yperpriors. Note that we make the dependence of $B_{N,K}$ on the hyperparameters Φ explicit in the notation of Proposition 4.3.6.

Proposition 4.3.6 (CRM truncation error with a hyperprior)**.** *Given hyperparameters* Φ*, consider a representation for* $\Theta \,|\, \Phi \sim \mathrm{CRM}(\nu)$*, and let* $B_{N,K}(\Phi)$ *be given by Eq.* (7) *(for a series representation) or Eq.* (17) *(for a superposition representation). The error of approximating* Θ *with its truncation* Θ_K *satisfies*

$$0 \leq \frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 \leq 1 - e^{-\mathbb{E}\left[B_{N,K}(\Phi)\right]} \leq 1.$$

Example 4.3.8 (Decoupled Bondesson representation truncation for $\Gamma\mathrm{P}(\gamma, \lambda, 0)$)**.** A standard choice of hyperprior for the mass γ is a gamma distribution, i.e. $\gamma \sim \mathsf{Gam}(a, b)$. Combining Proposition 4.3.6 and Example 4.3.5, we have that

$$\mathbb{E}\left[B_{N,K}(\Phi)\right] \leq N\frac{a}{b}\left(\frac{\xi}{\xi+1}\right)^{K}.$$

3.4 Normalized truncation analysis

In this section, we provide truncation error bounds for *normalized CRMs* (NCRMs). Examples include the Dirichlet process [Ferguson, 1973], the normalized gamma process [Brix, 1999, James, 2002, Lijoi and Prünster, 2003, Pitman, 2003, Lijoi et al., 2007, Lijoi and Prünster, 2010], and the normalized σ-stable process [Kingman, 1975, Lijoi and Prünster, 2010]. Given a CRM Θ on Ψ, we define t h e c o rresponding N C RM Ξ v i a $\Xi(S) := \Theta(S)/\Theta(\Psi)$ f o r e a ch measurable subset $S \subseteq \Psi$. Likewise, given a truncated CRM Θ_K, we define its normalization Ξ_K via

$\Xi_K(S) := \Theta_K(S)/\Theta_K(\Psi)$. Note that any simulation algorithm for Θ_K can be used for Ξ_K by simply normalizing the result. This does not depend on the particular representation of the CRM, and thus applies equally to all the representations in Section 4.2.

The first step in the analysis of NCRM truncations is to define their approximation error in a manner similar to that of CRM truncations. Since Ξ and Ξ_K are both normalized, they are distributions on Ψ; thus, observations $X_{1:N}$ are generated i.i.d. from Ξ, and $Z_{1:N}$ are generated i.i.d. from Ξ_K. $Y_{1:N}$ and $W_{1:N}$ have the same definition as for CRMs. As in the developments of Section 4.3, the theoretical results of this section rely on a general upper bound, provided by Lemma 4.4.1. Proposition 4.4.2 shows that the bound in Lemma 4.4.1 is tight without further assumptions on the data likelihood f.

Lemma 4.4.1 (NCRM protobound). *Let $\Theta \sim \mathrm{CRM}(\nu)$, and let its truncation be Θ_K. Let their normalizations be Ξ and Ξ_K respectively. If*

$$X_n \,|\, \Xi \overset{i.i.d.}{\sim} \Xi, \qquad\qquad Z_n \,|\, \Xi_K \overset{i.i.d.}{\sim} \Xi_K,$$
$$Y_n \,|\, X_n \overset{indep}{\sim} f(\cdot \,|\, X_n), \qquad\qquad W_n \,|\, Z_n \overset{indep}{\sim} f(\cdot \,|\, Z_n),$$

then

$$\frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 \leq 1 - \mathbb{P}\left(X_{1:N} \subseteq \mathrm{supp}(\Xi_K)\right),$$

where $p_{N,\infty}$, $p_{N,K}$ are the marginal densities of $Y_{1:N}$ and $W_{1:N}$, respectively.

Proposition 4.4.2 (Protobound tightness). *If G is non-atomic and Ψ is Borel, then for any $\delta > 0$, there exists a likelihood f such that Lemma 4.4.1 is tight up to a factor of $1 - \delta$.*

The analysis of CRMs in Section 4.3 relied heavily on the Poisson process stucture of the rates in Θ and $X_{1:N}$; unfortunately, the rates in Ξ do not possess the same structure and thus lack many useful independence properties (the rates must sum to one). Likewise, sampling X_n for each n does not depend on the atoms of Ξ independently (X_n randomly selects a single atom based on their rates). Rather than using the basic definitions of the above random quantities to derive an error bound, we decouple the atoms of Ξ and $X_{1:N}$ using a technique from extreme value theory. A *Gumbel* random variable T with location $\mu \in \mathbb{R}$ and scale $\sigma > 0$, denoted $T \sim \mathsf{Gumbel}(\mu, \sigma)$, is defined by the cumulative distribution function and corresponding density

$$\mathbb{P}(T \leq t) = e^{-e^{-\frac{t-\mu}{\sigma}}} \qquad \text{and} \qquad \frac{1}{\sigma}e^{-\left(\frac{t-\mu}{\sigma}\right) - e^{-\left(\frac{t-\mu}{\sigma}\right)}}.$$

An interesting property of the Gumbel distribution is that if one perturbs the log-probabilities of a finite discrete distribution by i.i.d. $\mathsf{Gumbel}(0, 1)$ random variables, the $\arg\max$ of the resulting set is a sample from the discrete distribution [Gumbel, 1954, Maddison et al., 2014]. This technique is invariant to normalization, as the $\arg\max$ is invariant to the corresponding constant shift in the log-transformed space. For present purposes, we develop the infinite extension of this result:

Lemma 4.4.3 (Infinite Gumbel-max sampling). *Let $(p_i)_{i=1}^\infty$ be a collection of positive numbers such that $\sum_i p_i < \infty$ and let $\bar{p}_j := \frac{p_j}{\sum_i p_i}$. If $(T_i)_{i=1}^\infty$ are i.i.d. $\mathsf{Gumbel}(0, 1)$ random variables, then $\arg\max_{i \in \mathbb{N}} T_i + \log p_i$ exists, is unique a.s., and has distribution*

$$\arg\max_{i \in \mathbb{N}} \; T_i + \log p_i \sim \mathsf{Categorical}\left((\bar{p}_j)_{j=1}^\infty\right).$$

21

The proof of this result, along with the others in this section, may be found in Appendix A.4. The utility of Lemma 4.4.3 is that it allows the construction of Ξ and $X_{1:N}$ without the problematic coupling of the underlying CRM atoms due to normalization; rather than dealing directly with Ξ, we log-transform the rates of Θ, perturb them by i.i.d. $\mathsf{Gumbel}(0,1)$ random variables, and characterize the distribution of the maximum rate in this process. The combination of this distribution with Lemma 4.4.3 yields the key proof technique used to develop the truncation bounds in Theorems 4.4.4 and 4.4.5. The results presented in this section are summarized in Table 1 in Chapter 5.

3.4.1 Series representations

The following result provides a general truncation error bound for normalized series representations, specifies its range, and guarantees that it decays to 0 as $K \to \infty$. We again use the general series representation notation from Eq. (6), where g is a distribution on $\mathbb{R}_+$, and $\tau : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is a measurable function such that $\lim_{u \to \infty} \tau(v, u) = 0$ for g-almost every v.

Theorem 4.4.4 (Normalized series representation truncation error bound). *The error of approximating a series representation of* $\Xi \sim \mathrm{NCRM}(\nu)$ *with its truncation* Ξ_K *satisfies*

$$0 \leq \frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 \leq 1 - (1 - B_K)^N \leq 1,$$

where

$$B_K := \mathbb{E}\left[\int_0^\infty J\left(\Gamma_K, t\right)\left(\int_0^1 J\left(\Gamma_K u, t\right)\, \mathrm{d}u\right)^{K-1}\left(-\frac{\mathrm{d}}{\mathrm{d}t}e^{\int_0^\infty (J(u+\Gamma_K, t)-1)\mathrm{d}u}\right)\mathrm{d}t\right], \qquad (20)$$

$$J(u,t) = \mathbb{E}\left[e^{-t\cdot\tau(V,u)}\right], \qquad V \sim g, \qquad and \qquad \Gamma_K \sim \mathsf{Gam}(K,1).$$

Furthermore, $\lim_{K \to \infty} B_K = 0$.

Example 4.4.1 (Dirichlet process, $\mathrm{DP}(\gamma)$, B-Rep). The Dirichlet process with concentration $\gamma > 0$ is a normalized gamma process $\mathrm{N\Gamma P}(\gamma, 1, 0)$. From Example 4.2.2 we have $c_\nu = \gamma$ and $g_\nu = \mathsf{Exp}(1)$, and from Section 4.3.1 we have $\tau(v, u) = ve^{-u/c_\nu}$. Therefore J and its antiderivative are

$$J(u,t) = \mathbb{E}\left[e^{-tVe^{-u/\gamma}}\right] = \left(1 + te^{-u/\gamma}\right)^{-1} \quad \text{and} \quad \int J(u,t)\mathrm{d}u = \gamma \log\left(e^{u/\gamma} + t\right).$$

Using the antiderivative to evaluate the integrals in the formula for B_K, writing the expectation over $\Gamma_K \sim \mathsf{Gam}(K,1)$ explicitly, and making a change of variables we have

$$B_K = \frac{\gamma^{K+1}}{\Gamma(K)}\int_0^\infty \int_1^\infty \left(\log\left(\frac{s+t}{1+t}\right)\right)^{K-1}(s+t)^{-(\gamma+2)}\, \mathrm{d}s\, \mathrm{d}t = \left(\frac{\gamma}{1+\gamma}\right)^K,$$

where the last equality is found by multiplying and dividing the integrand by $(1+t)^{-(\gamma+2)}$, and making the change of variables from s to $x = \log\frac{s+t}{1+t}$. Therefore, the truncation error can be bounded by

$$\frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 \leq 1 - \left(1 - \left(\frac{\gamma}{\gamma+1}\right)^K\right)^N \sim N\left(\frac{\gamma}{\gamma+1}\right)^K \qquad K \to \infty.$$

The bound in Example 4.4.2 has exponential decay, and reproduces earlier DP truncation error bound rates due to Ishwaran and James [2001] and Ishwaran and Zarepour [2002]. However, the techniques used in past work do not generalize beyond the Dirichlet process, while those developed here apply to any NCRM.

4.4.2 Superposition representations

The following result provides a general truncation error bound for normalized superposition representations, specifies its range, and guarantees that it decays to 0 as $K \to \infty$. We once again rely on the property that the truncation Θ_K and tail Θ_K^+ are mutually independent CRMs, as expressed in Eq. (16), with the tail measure denoted ν_K^+.

Theorem 4.4.5 (Truncation error bound for normalized superposition representations)**.** *The error of approximating a superposition representation of $\Xi \sim \mathrm{NCRM}(\nu)$ with its truncation Ξ_K satisfies*

$$0 \leq \frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 \leq 1 - (1 - B_K)^N \leq 1,$$

where

$$B_K := \int_0^\infty \left(\int \theta e^{-\theta t} \nu_K^+(\mathrm{d}\theta) \right) e^{\int \left(e^{-\theta t} - 1 \right) \nu(\mathrm{d}\theta)} \, \mathrm{d}t. \tag{21}$$

Furthermore, $\lim_{K \to \infty} B_K = 0$.

This bound can be applied by using the tail measures derived earlier in Section 4.3.2.

Example 4.4.2 (Dirichlet process, $\mathrm{DP}(\gamma)$, DB-Rep)**.** As in Example 4.4.1, we view the Dirichlet process with concentration $\gamma > 0$ as a normalized gamma process $\mathrm{N\Gamma P}(\gamma, 1, 0)$. First, by Lemma A.1.8, the integral in the exponential is

$$\exp\left(\int (e^{-t\theta} - 1)\nu(\mathrm{d}\theta) \right) = \exp\left(\gamma \int_0^\infty (e^{-t\theta} - 1)\theta^{-1}e^{-\theta}\mathrm{d}\theta \right) = (t+1)^{-\gamma}.$$

Example 4.2.2 shows $c_\nu = \gamma$ and $g_\nu(v) = e^{-v}$, and Eq. (27) provides the tail measure ν_K^+ for the decoupled Bondesson representation,

$$\nu_K^+(\mathrm{d}\theta) = \frac{\gamma}{\xi} \sum_{k=K+1}^\infty \frac{\xi^k}{\Gamma(k)} \left(\int_0^1 (-\log x)^{k-1} x^{\xi-2} e^{-\theta x^{-1}} \mathrm{d}x \right) \mathrm{d}\theta.$$

Substituting this result, using Fubini's theorem to swap the order of integration and summation, evaluating the integral over θ, and making the substitution $x = e^{-s}$ yields

$$B_K = \frac{\gamma}{\xi} \sum_{k=K+1}^\infty \frac{\xi^k}{\Gamma(k)} \iint_{s,t \geq 0} \frac{s^{k-1} e^{-(\xi-1)s}(t+1)^{-\gamma}}{(e^s + t)^2} \mathrm{d}s \, \mathrm{d}t.$$

Noting that $\forall s \geq 0$, $e^s \geq 1$, we have for any $a \in (0,1] \cap (0,\gamma)$,

$$B_K \leq \frac{\gamma}{\xi} \sum_{k=K+1}^\infty \frac{\xi^k}{\Gamma(k)} \int_0^\infty s^{k-1} e^{-(\xi+a)s}\mathrm{d}s \int_0^\infty (t+1)^{-(\gamma+1-a)}\mathrm{d}t$$

$$= \frac{\gamma}{(\gamma-a)\xi} \sum_{k=K+1}^\infty \left(\frac{\xi}{\xi+a} \right)^k = \frac{\gamma}{a(\gamma-a)} \left(\frac{\xi}{\xi+a} \right)^K.$$

23

Therefore, for any $a \in (0, 1] \cap (0, \gamma)$,

$$\frac{1}{2} \|p_{N,\infty} - p_{N,K}\|_1 \leq 1 - \left(1 - \frac{\gamma}{a(\gamma - a)} \left(\frac{\xi}{\xi + a}\right)^K\right)^N \sim \frac{N\gamma}{a(\gamma - a)} \left(\frac{\xi}{\xi + a}\right)^K \quad K \to \infty.$$

To find the tightest bound, one can minimize with respect to a given γ, ξ, K.

Example 4.4.3 (Normalized gamma process, $\mathrm{N\Gamma P}(\gamma, \lambda, d)$, SB-Rep). By Lemma A.1.8, the integral in the exponential is

$$\exp\left(\int (e^{-\theta t} - 1)\nu(\mathrm{d}\theta)\right) = \begin{cases} \exp\left(-\gamma\lambda^{1-d}d^{-1}((t + \lambda)^d - \lambda^d)\right) & d > 0 \\ \left(\frac{\lambda}{t+\lambda}\right)^{\gamma\lambda} & d = 0, \end{cases} \tag{22}$$

and the standard gamma integral yields

$$\int \theta e^{-\theta t} \nu_K^+(\mathrm{d}\theta) = \gamma \frac{\lambda^{1-d}}{\Gamma(1-d)} \int \theta^{-d} e^{-(K+t+\lambda)\theta} \, \mathrm{d}\theta = \gamma\lambda^{1-d}(K + t + \lambda)^{d-1}. \tag{23}$$

When $d > 0$, multiplying the previous two displays and integrating over $t \geq 0$ yields

$$B_K = \gamma\lambda^{1-d}e^{\gamma\lambda/d} \int_\lambda^\infty (K + t)^{d-1} e^{-\gamma\lambda^{1-d}t^d/d} \, \mathrm{d}t \leq C_{\gamma,\lambda,d} \left(K + \lambda\right)^{d-1},$$

where we have used $(K + t)^{d-1} \leq (K + \lambda)^{d-1}$ for $t \geq \lambda$ and the change of variables $u = \gamma\lambda^{1-d}d^{-1}t^d$ to find that $C_{\gamma,\lambda,d} = e^\sigma \sigma^{1-d^{-1}}\lambda^{1-d}\Gamma(d^{-1}, \sigma)$, where $\sigma = \gamma\lambda d^{-1}$ and $\Gamma(a, x) := \int_x^\infty \theta^{a-1}e^{-\theta}\mathrm{d}\theta$ is the upper incomplete gamma function. Therefore,

$$\frac{1}{2} \|p_{N,\infty} - p_{N,K}\|_1 \leq 1 - \left(1 - C_{\gamma,\lambda,d}(K + \lambda)^{d-1}\right)^N \sim NC_{\gamma,\lambda,d}K^{d-1} \quad K \to \infty.$$

When $d = 0$, multiplying Eqs. (22) and (23) and integrating over $t \geq 0$ yields

$$\frac{B_K}{\gamma\lambda^{1+\gamma\lambda}} = \int_\lambda^\infty (K + t)^{-1}t^{-\gamma\lambda}\mathrm{d}t \leq \begin{cases} (K + \lambda)^{-1} \left(\frac{\frac{1}{\gamma\lambda}(K+\lambda)^{1-\gamma\lambda} - \lambda^{1-\gamma\lambda}}{1-\gamma\lambda}\right) & \gamma\lambda \neq 1 \\ K^{-1}\log\left(\frac{K+\lambda}{\lambda}\right) & \gamma\lambda = 1, \end{cases}$$

where we obtain the bound for $\gamma\lambda \neq 1$ by splitting the integral into the intervals $[\lambda, K + \lambda]$ and $[K + \lambda, \infty)$ and bounding each section separately, and we obtain the bound for $\gamma\lambda = 1$ via the transformation $u = t/(K + t)$. Therefore, asymptotically

$$\frac{1}{2} \|p_{N,\infty} - p_{N,K}\|_1 \lesssim N \begin{cases} C_{\gamma,\lambda}K^{-\min(1,\gamma\lambda)} & \gamma\lambda \neq 1 \\ \lambda K^{-1}\log K & \gamma\lambda = 1 \end{cases} \quad K \to \infty,$$

where $C_{\gamma,\lambda} := \max\left(\frac{\lambda^{\gamma\lambda}}{1-\gamma\lambda}, \frac{\gamma\lambda^2}{\gamma\lambda-1}\right)$.

Truncation of the $\mathrm{N\Gamma P}(\gamma, \lambda, d)$ has been studied previously: Argiento et al. [2016] threshold the weights of the unnormalized CRM to be beyond a fixed level $\epsilon > 0$ prior to normalization, and develop error bounds for that method of truncation. These results are not directly comparable to those of the present work due to the different methods of truncation (i.e. sequential representation termination versus weight thresholding).

24

3.4.3 Hyperpriors

As in the CRM case, we can place priors on the hyperparameters of the NCRM rate measure (i.e. γ, α, λ, d, etc.). We conclude our investigation of NCRM truncation error by showing how bounds developed in this section can be modified to account for hyperpriors. Note that we make the dependence of B_K on the hyperparameters Φ explicit in the notation of Proposition 4.4.6.

Proposition 4.4.6 (NCRM truncation error with a hyperprior). *Given hyperparameters Φ, consider a representation for $\Theta \,|\, \Phi \sim \mathrm{CRM}(\nu)$, let $\Xi \,|\, \Phi$ be its normalization, and let $B_K(\Phi)$ be given by Eq. (20) (for a series representation) or Eq. (21) (for a superposition representation). The error of approximating Ξ with its truncation Ξ_K satisfies*

$$0 \le \frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 \le 1 - (1 - \mathbb{E}\left[B_K(\Phi)\right])^N \le 1.$$

Example 4.4.4 (Dirichlet process, $\mathrm{DP}(\gamma)$, B-Rep). If we place a Lomax prior on γ, i.e. $\gamma \sim \mathrm{LomP}(a, 1)$, then combining Proposition 4.4.6 and Example 4.4.1 yields

$$\frac{1}{2}\left\|p_{N,\infty} - p_{N,K}\right\|_1 \le 1 - \left(1 - \frac{\Gamma(a+1)\Gamma(K+1)}{\Gamma(a+K+1)}\right)^N \sim N\Gamma(a+1)(K+1)^{-a} \quad K \to \infty.$$

3.5 Simulation and computational complexity

The sequential representations in Section 4.2 are each generated from a different finite sequence of distributions, resulting in a different expected computational cost for the same truncation level. Thus, the truncation level itself is not an appropriate parameter with which to compare the error bounds for different representations and we require a characterization of the computational cost. We investigate the mean complexity $\mathbb{E}[R]$ of each representation, where R is the number of random variables sampled, as a function of the truncation level for each of the representations in Section 4.2.

We begin with the series representations. For each value of $k = 1, \ldots, K$, each series representation generates a single trait $\psi_k \sim G$ and a rate θ_k composed of some transformation of random variables. Thus, all of the series representations in this work satisfy $\mathbb{E}[R] = rK$ for some constant r: by inspection, the inverse-Lévy representation has $r = 2$, and all the remaining series representations have $r = 3$.

The superposition representations, on the other hand, generate a Poisson random variable to determine the number of atoms at each value of $k = 1, \ldots, K$, and then generate those atoms. Therefore, the mean simulation complexity takes the form $\mathbb{E}[R] = \sum_{k=1}^{K} 1 + r_k \mathbb{E}[C_k]$ for some constants r_k that might depend on the value of k. For the decoupled Bondesson representation, $r_k = 3$ since each atom requires generating three values (ψ_{ki}, V_{ki}, and T_{ki}), and $\mathbb{E}[C_k] = c/\xi$, so $\mathbb{E}[R] = \left(\frac{3c}{\xi} + 1\right)K$. For the size-biased representation, $r_k = 3$ since each atom requires generating three values (ψ_{ki}, z_{ki}, and θ_{ki}), and $\mathbb{E}[C_k] = \eta_k$, so $\mathbb{E}[R] = K + 3\sum_{k=1}^{K} \eta_k$. Note that here $\mathbb{E}[R] \sim K$, for $K \to \infty$ since η_k is a decreasing sequence. For the power-law representation, $r_k = k+2$, since each atom requires generating ψ_{ki}, V_{ki}, and k beta random variables, and therefore $\mathbb{E}[R] = \left(1 + \frac{5\gamma}{2}\right)K + \frac{\gamma}{2}K^2$.
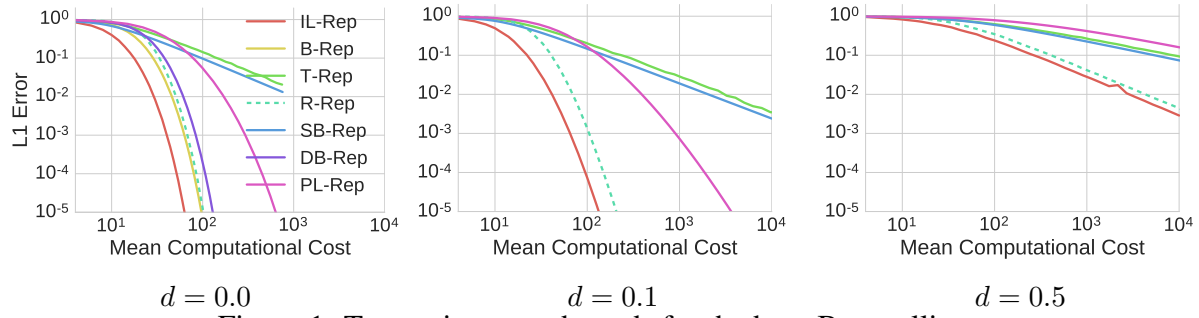
Figure 1: Truncation error bounds for the beta-Bernoulli process.

3.6 Additional Examples

In this section, we provide example applications of our theory to the beta process and to the beta prime process.

3.6.1 Beta process

The beta process $\mathrm{BP}(\gamma, \alpha, d)$ [Teh and Görür, 2009, Broderick et al., 2012] with discount parameter $d \in [0, 1)$, concentration parameter $\alpha > -d$, and mass parameter $\gamma > 0$, is a CRM with rate measure

$$\nu(\mathrm{d}\theta) = \gamma \frac{\Gamma(\alpha + 1)}{\Gamma(1 - d)\Gamma(\alpha + d)} \mathbb{1}\left[\theta \leq 1\right] \theta^{-1-d}(1 - \theta)^{\alpha+d-1}\mathrm{d}\theta. \tag{24}$$

Setting $d = 0$ yields the standard beta process [Hjort, 1990, Thibaux and Jordan, 2007]. The beta process is often paired with a Bernoulli likelihood or negative binomial likelihood with $s \in \mathbb{N}$ failures:

$$\text{Bern:} \qquad h(x \mid \theta) = \mathbb{1}\left[x \leq 1\right]\theta^x(1 - \theta)^{1-x},$$

$$\text{NegBinom:} \qquad h(x \mid \theta) = \binom{x + s - 1}{x}(1 - \theta)^s\theta^x.$$

Note that for the Bernoulli likelihood $\pi(\theta) = 1 - \theta$ and for the negative binomial likelihood $\pi(\theta) = (1 - \theta)^s$.

Bondesson representation If $\alpha > 1$ and $d = 0$, then $\theta\nu(\theta) = \gamma\alpha(1 - \theta)^{\alpha-1}\mathbb{1}[\theta \leq 1]$ is non-increasing, $c_\nu = \lim_{\theta \to 0} \theta\nu(\theta) = \gamma\alpha$, and $g_\nu(v) = (\alpha - 1)(1 - v)^{\alpha-2} = \mathsf{Beta}(v; 1, \alpha - 1)$. Thus, it follows from Theorem 4.2.1 that if $\Theta \leftarrow \mathrm{B\text{-}Rep}(\gamma\alpha, \mathsf{Beta}(1, \alpha - 1))$, then $\Theta \sim \mathrm{BP}(\gamma, \alpha, 0)$. In the case of $\alpha = 1$, $g_\nu(v) = \delta_1$, so $V_k \equiv 1$ and the Bondesson representation is equivalent to the inverse-Lévy representation. Since $\exp(-E_k/c) \sim \mathsf{Beta}(1, c)$, the representation used in Teh et al. [2007] is equivalent to the Bondesson representation for $\mathrm{BP}(\gamma, 1, 0)$.

26

To obtain a truncation bound in the Bernoulli likelihood case, we can argue as in Example 4.3.2:

$$B_{N,K} \leq N \int_0^\infty \left(1 - \mathbb{E}\left[\pi(ve^{-G_K/(\gamma\alpha)})\right]\right) \nu(dv)$$

$$= N\gamma\alpha \, \mathbb{E}[e^{-G_K/(\gamma\alpha)}] \int_0^1 (1-v)^{\alpha-1} dv \qquad (25)$$

$$= N\gamma \left(\frac{\gamma\alpha}{1+\gamma\alpha}\right)^K.$$

This result generalizes that in Doshi-Velez et al. [2009], which applies only when $\alpha = 1$.

Theorem 4.2.1 does not apply directly when $\alpha < 1$, since $\lim_{\theta \to 0} \theta\nu(\theta) = \infty$. However, a representation can be obtained by using a trick from Paisley et al. [2012]. For $\alpha > 0$, let

$$\Theta' = \sum_{k=1}^C \theta'_k \delta_{\psi'_k},$$

where $C \sim \mathsf{Poiss}(\gamma)$, $\theta'_k \overset{\text{i.i.d.}}{\sim} \mathsf{Beta}(1, \alpha)$, and $\psi'_k \overset{\text{i.i.d.}}{\sim} G$. Thus, Θ' is a CRM with rate measure $\gamma\alpha(1-\theta)^{\alpha-1}\mathbb{1}[\theta \leq 1]d\theta$. If $\Theta \sim \mathrm{BP}(\gamma\alpha/(\alpha+1), \alpha+1, 0)$, which can be generated according to Theorem 4.2.1, then $\Theta'' = \Theta + \Theta'$ is a CRM with rate density on $[0, 1]$ given by

$$\gamma\alpha\theta^{-1}(1-\theta)^\alpha + \gamma\alpha(1-\theta)^{\alpha-1} = \gamma\alpha\theta^{-1}(1-\theta)^{\alpha-1},$$

hence $\Theta'' \sim \mathrm{BP}(\gamma, \alpha, 0)$.

Thinning representation   If we let $g = \mathsf{Beta}(1-d, \alpha+d)$, then the thinning representation for $\mathrm{BP}(\gamma, \lambda, d)$ is

$$\Theta = \sum_{k=1}^\infty V_k \mathbb{1}(V_k\Gamma_k \leq \gamma)\delta_{\psi_k}, \qquad \text{with} \qquad V_k \overset{\text{i.i.d.}}{\sim} \mathsf{Beta}(1-d, \alpha+d).$$

Rejection representation   To obtain a rejection representation for any d when $\alpha \geq 1-d$, let μ be the rate measure for $\mathrm{BP}(\gamma\frac{\Gamma(\alpha+1)}{\Gamma(2-d)\Gamma(\alpha+d)}, 1-d, d)$. We then have that

$$\mu[x, \infty) = \begin{cases} \gamma'd^{-1}(x^{-d}-1) & d > 0 \\ -\gamma'\log x & d = 0 \end{cases} \quad \text{and} \quad \mu^{\leftarrow}(u) = \begin{cases} (1+du/\gamma')^{-1/d} & d > 0 \\ e^{-u/\gamma'} & d = 0, \end{cases}$$

where $\gamma' := \gamma\frac{\Gamma(\alpha+1)}{\Gamma(1-d)\Gamma(\alpha+d)}$. Thus, we can apply the inverse-Lévy method analytically for μ. Since we have constructed μ such that $d\nu/d\mu \leq 1$, we can use μ to construct the rejection representation

$$\Theta = \sum_{k=1}^\infty V_k \mathbb{1}(U_k \leq (1-V_k)^{\alpha+d-1}) \sim \mathrm{BP}(\gamma, \alpha, d), \qquad \alpha \geq 1-d,$$

$$\text{with} \qquad V_k = \begin{cases} (1+d\Gamma_k/\gamma')^{-1/d} & d > 0 \\ e^{-\Gamma_k/\gamma'} & d = 0 \end{cases}, \qquad U_k \overset{\text{i.i.d.}}{\sim} \mathsf{Unif}[0, 1].$$

27

The expected number of rejections is

$$-d^{-1}\left[1 + 2d\,_2F_1^{(0,0,1,0)}(-\alpha - d, -d, -d; 1) + d\,_2F_1^{(0,1,0,0)}(-\alpha - d, -d, -d; 1)\right],$$

where $_2F_1$ is the ordinary hypergeometric function and the parenthetical superscripts indicate partial derivatives. This quantity monotonically diverges to ∞ as $d \to 1$.

To obtain a truncation bound in the Bernoulli likelihood case, we consider the $d > 0$ and $d = 0$ settings separately. If $d > 0$, we have

$$\begin{aligned}
\frac{B_{N,K}}{N\gamma'} &\leq \int_0^1 F_K(\gamma'd^{-1}(x^{-d} - 1))x^{-d}(1-x)^{\alpha+d-1}\mathrm{d}x \\
&\leq \int_0^a x^{-d}(1-x)^{a+d-1}\mathrm{d}x + F_K(\gamma'd^{-1}(a^{-d} - 1))\int_a^1 x^{-d}(1-x)^{\alpha+d-1}\mathrm{d}x \\
&\leq \int_0^a x^{-d}\mathrm{d}x + a^{-d}F_K(\gamma'd^{-1}(a^{-d} - 1))\int_a^1 (1-x)^{\alpha+d-1}\mathrm{d}x \\
&\leq (1-d)^{-1}a^{1-d} + a^{-d}\left(3\gamma'd^{-1}(a^{-d} - 1)/K\right)^K \\
&\leq (1-d)^{-1}a^{1-d} + \left(3\gamma'd^{-1}/K\right)^K a^{-(K+1)d}.
\end{aligned}$$

Setting the two terms equal and solving for a we obtain $a^{1+d(K-1)} = (3\gamma'(d^{-1} - 1)/K)^K$ and conclude that

$$B_{N,K} \leq 2N\gamma'\left(\frac{3\gamma'(d^{-1} - 1)}{K}\right)^{\frac{K}{1+d(K-1)}} \sim 2N\gamma'\left(\frac{3\gamma'(d^{-1} - 1)}{K}\right)^{1/d} \qquad K \to \infty.$$

If $d = 0$, we have

$$\begin{aligned}
\frac{B_{N,K}}{N\gamma'} &\leq \int_0^1 F_K(-\gamma'\log x)(1-x)^{\alpha-1}\mathrm{d}x \\
&\leq \int_0^a (1-x)^{\alpha-1}\mathrm{d}x + F_K(-\gamma'\log a)\int_a^1 (1-x)^{\alpha-1}\mathrm{d}x \\
&\leq a + (-3\gamma'\log a/K)^K.
\end{aligned}$$

Setting the two terms equal and solving for a we conclude that

$$B_{N,K} \leq 2N\gamma\alpha e^{-KW_0(\{3\gamma\alpha\}^{-1})},$$

where W_0 is as defined in Eq. (11).

Decoupled Bondesson and power-law representations The decoupled Bondesson representation for $\mathrm{BP}(\gamma, \alpha, 0)$ from Paisley et al. [2010] was extended by Broderick et al. [2012] to the $\mathrm{BP}(\gamma, \alpha, d)$ setting. The Broderick et al. [2012] construction for the $\mathrm{BP}(\gamma, \alpha, d)$ is in fact the "trivial" power-law representation $\mathrm{PL\text{-}Rep}(\gamma, \alpha, d, \delta_1)$ (the decoupled Bondesson representation is the special case when $d = 0$).

In the Bernoulli likelihood case, the truncation bound for the decoupled Bondesson representation is

$$B_{N,K} \leq \frac{N\gamma\alpha}{\xi}\sum_{k=K+1}^{\infty}\mathbb{E}[Ve^{-T_k}] = \frac{N\gamma\alpha}{\xi}\sum_{k=K+1}^{\infty}\frac{1}{\alpha}\left(\frac{\xi}{1+\xi}\right)^K = N\gamma\left(\frac{\xi}{1+\xi}\right)^K.$$

28

For the power law representation, by the same arguments as Example 4.3.7,

$$B_{N,K} \le N\gamma \prod_{k=1}^{K} \frac{\alpha + kd}{\alpha + kd - d + 1}.$$

This result generalizes that in Paisley et al. [2012], which applies only when $d = 0$.

Size-biased representation The size-biased representation of the beta process is well-established and we refer the reader to Broderick et al. [2012, 2017] for details. We note that the standard beta integral yields

$$\eta_k = \eta_{k1} = \int \pi(\theta)^{k-1}(1 - \pi(\theta))\nu(\mathrm{d}\theta) = \gamma\frac{\Gamma(\alpha + 1)}{\Gamma(\alpha + d)}\frac{\Gamma(\alpha + d + k - 1)}{\Gamma(\alpha + k)},$$

and for $i > 1$, $\eta_{ki} = 0$. Hence $z_{ki} = 1$ almost surely and $\theta_{ki} \sim \mathsf{Beta}(1 - d, \alpha + d + k - 1)$, demonstrating that the construction due to Thibaux and Jordan [2007] is a special case of the size-biased representation for $\mathrm{BP}(\gamma, \alpha, d)$.

To obtain a truncation bound for the Bernoulli likelihood case, first consider the $d > 0$ setting. Using Lemma A.1.6 to simplify the sum in Eq. (18), we have

$$B_{N,K} \le \frac{\gamma}{d}\frac{\Gamma(\alpha + 1)}{\Gamma(\alpha + d)}\left(\frac{\Gamma(\alpha + d + K)}{\Gamma(\alpha + K)} - \frac{\Gamma(\alpha + d + K + N)}{\Gamma(\alpha + K + N)}\right)$$
$$\sim \gamma N\frac{\Gamma(\alpha + 1)}{\Gamma(\alpha + d)}K^{d-1} \qquad K \to \infty,$$

where the asymptotic result follows from Lemmas A.1.4, A.1.7 and A.1.9. When $d = 0$, we can again use Lemma A.1.6 to arrive at

$$B_{N,K} \le \gamma\alpha\left(\psi(\alpha + K) - \psi(\alpha + N + K)\right) \sim \gamma\alpha N K^{-1} \qquad K \to \infty,$$

where $\psi(\cdot)$ is the digamma function, and the asymptotic result follows from Lemma A.1.4.

We can also bound the truncation error in the case of the negative binomial likelihood. For a fixed number of failures $s > 0$, and assuming $\alpha + d + (k - 1)s > 1$, integration by parts yields

$$\int \pi(\theta)^{k-1}(1 - \pi(\theta))\nu(\mathrm{d}\theta)$$
$$= \frac{\gamma}{d}\frac{\Gamma(\alpha + 1)}{\Gamma(\alpha + d)}\left(\frac{\Gamma(\alpha + d + ks)}{\Gamma(\alpha + ks)} - \frac{\Gamma(\alpha + d + (k - 1)s)}{\Gamma(\alpha + (k - 1)s)}\right), \qquad (26)$$

When $d > 0$, the sum from Eq. (18) is telescoping, so canceling terms,

$$B_{N,K} \le \frac{\gamma}{d}\frac{\Gamma(\alpha + 1)}{\Gamma(\alpha + d)}\left(\frac{\Gamma(\alpha + d + Ks)}{\Gamma(\alpha + Ks)} - \frac{\Gamma(\alpha + d + (K + N)s)}{\Gamma(\alpha + (K + N)s)}\right)$$
$$\sim \gamma N s^d\frac{\Gamma(\alpha + 1)}{\Gamma(\alpha + d)}K^{d-1} \qquad K \to \infty,$$

where the asymptotic result follows from Lemmas A.1.4, A.1.7 and A.1.9. To analyze the case where $d = 0$, we can use L'Hospital's rule to take the limit of Eq. (26) as $d \to 0$, yielding

$$\lim_{d \to 0} \int \pi(\theta)^{k-1}(1 - \pi(\theta))\nu(\mathrm{d}\theta) = \gamma\alpha(\psi(\alpha + ks) - \psi(\alpha + (k - 1)s)).$$

29

Again computing the error bound by canceling terms in the telescoping sum,

$$B_{N,K} \leq \gamma\alpha \left(\psi(\alpha + Ks) - \psi(\alpha + (K+N)s)\right) \sim \gamma\alpha N K^{-1} \qquad K \to \infty,$$

where the asymptotic result follows from an application of Lemma A.1.4.

Stochastic mapping We can transform the gamma process $\Gamma\mathrm{P}(\gamma, \lambda, 0)$ into the beta process $\mathrm{BP}(\gamma, \alpha, 0)$ by applying the stochastic mapping

$$\theta \mapsto \theta/(\theta + G), \qquad G \sim \mathsf{Gam}(\alpha, \alpha), \qquad \kappa(\theta, \mathrm{d}u) = \mathsf{Gam}(\theta/(\theta + u); \alpha, \alpha) \frac{\theta}{u^2} \mathrm{d}u.$$

Using Lemma 4.2.4 yields $\kappa(\Theta) \sim \mathrm{BP}(\gamma, \alpha, 0)$. Applying this result to the Bondesson representation for $\Gamma\mathrm{P}(\gamma, \alpha, 0)$ yields

$$\sum_{k=1}^{\infty} \theta_k \delta_{\psi_k} \sim \mathrm{BP}(\gamma, \alpha, 0), \quad \text{with} \qquad \theta_k := (1 + G_k V_k^{-1} e^{\Gamma_{k,\alpha\gamma}})^{-1}, \qquad \psi_k \overset{\text{i.i.d.}}{\sim} G,$$

$$G_k \overset{\text{i.i.d.}}{\sim} \mathsf{Gam}(\alpha, \alpha), \qquad\qquad V_k \overset{\text{i.i.d.}}{\sim} \mathsf{Exp}(\alpha).$$

which, unlike the Bondesson representation, applies for all $\alpha > 0$.

Hyperpriors Consider truncating the Bondesson representation of the beta process, but with a hyperprior on the mass parameter γ. A standard choice of hyperprior for γ is a gamma distribution, i.e. $\gamma \sim \mathsf{Gam}(a, b)$. Combining Proposition 4.3.6 and the beta-Bernoulli truncation bound in Eq. (25), we have that

$$B_{N,K} \leq N \frac{a}{b} \left(\frac{\xi}{\xi + 1}\right)^K.$$

3.6.2 Beta prime process

The beta prime process $\mathrm{BPP}(\gamma, \alpha, d)$ [Broderick et al., 2017] with discount parameter $d \in [0, 1)$, concentration parameter $\alpha > -d$, and mass parameter $\gamma > 0$, is a CRM with rate measure

$$\nu(\mathrm{d}\theta) = \gamma \frac{\Gamma(\alpha + 1)}{\Gamma(1 - d)\Gamma(\alpha + d)} \theta^{-1-d}(1 + \theta)^{-\alpha} \mathrm{d}\theta.$$

The beta prime process is often paired with an odds Bernoulli likelihood,

$$h(x \mid \theta) = \mathbb{1}\left[x \leq 1\right] \theta^x (1 + \theta)^{-1},$$

in which case $\pi(\theta) = (1 + \theta)^{-1}$. All truncation results are for the odds Bernoulli likelihood.

Bondesson representation If $d = 0$, then $\theta\nu(\theta) = \gamma\alpha(1 + \theta)^{-\alpha}$ is non-increasing and $c_\nu = \lim_{\theta \to 0} \theta\nu(\theta) = \gamma\alpha$, so $g_\nu(v) = \alpha(1 + v)^{-\alpha-1} = \mathsf{Beta}'(v; 1, \alpha)$. Thus, it follows from Theorem 4.2.1 that if $\Theta \leftarrow \mathrm{B\text{-}Rep}(\gamma\alpha, \mathsf{Beta}'(1, \alpha))$, then $\Theta \sim \mathrm{BPP}(\gamma, \alpha, 0)$. For the truncation
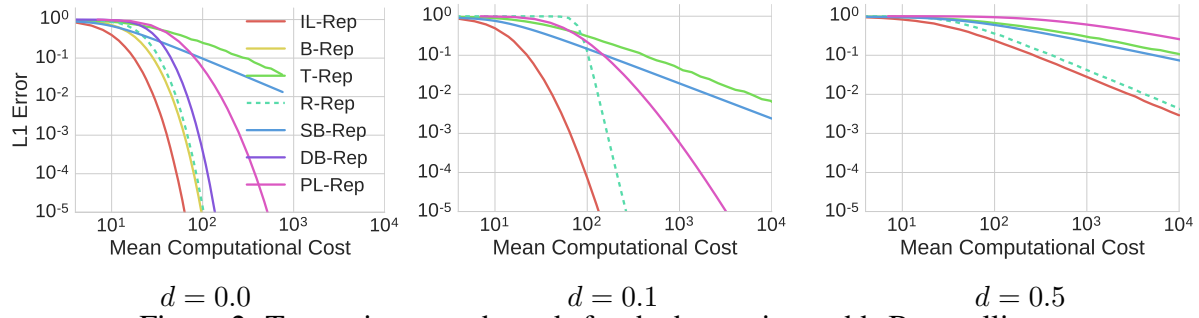
Figure 2: Truncation error bounds for the beta prime-odds Bernoulli process.

bound we have

$$
\begin{aligned}
B_{N,K} &= N \int_0^\infty \left(1 - \mathbb{E}\left[\pi(ve^{-G_K/(\gamma\alpha)})\right]\right) \nu(\mathrm{d}v) \\
&= N\gamma\alpha\, \mathbb{E}\left[e^{-G_K/(\gamma\alpha)} \int_0^\infty (1+v)^{-\alpha}(1+ve^{-G_K/(\gamma\alpha)})^{-1}\mathrm{d}v\right] \\
&\leq N\gamma\alpha\, \mathbb{E}[e^{-G_K/(\gamma\alpha)}] \int_0^\infty (1+v)^{-\alpha}(1+v\mathbb{E}[e^{-G_K/(\gamma\alpha)}])^{-1}\mathrm{d}v \\
&\leq N\gamma\alpha\left(\frac{\gamma\alpha}{1+\gamma\alpha}\right)^K \int_0^\infty (1+v)^{-\alpha-1}\mathrm{d}v \\
&= N\gamma\left(\frac{\gamma\alpha}{1+\gamma\alpha}\right)^K,
\end{aligned}
$$

where the first upper bound follows from Jensen's inequality. Thus, the error bound is the same as for the beta-Bernoulli process.

Thinning representation If we let $g = \mathsf{Beta}'(1-d, \alpha+d)$, then the thinning representation for $\mathrm{BPP}(\gamma, \alpha, d)$ is

$$
\Theta = \sum_{k=1}^\infty V_k \mathbb{1}(V_k(\Gamma_k - \gamma) \leq \gamma)\delta_{\psi_k}, \qquad \text{with} \qquad V_k \overset{\text{i.i.d.}}{\sim} \mathsf{Beta}'(1-d, \alpha+d).
$$

Rejection representation For $d = 0$ and $\alpha \geq 1$, we take μ to be the rate measure for $\mathrm{LomP}(\gamma, 1)$, so the rejection representation is

$$
\Theta = \sum_{k=1}^\infty V_k \mathbb{1}(U_k \leq (1+V_k)^{\alpha-1}) \sim \mathrm{BPP}(\gamma, \alpha, 0), \qquad \alpha \geq 1,
$$

$$
\text{with} \qquad V_k = (e^{\gamma^{-1}\alpha^{-1}\Gamma_k} - 1)^{-1}, \qquad U_k \overset{\text{i.i.d.}}{\sim} \mathsf{Unif}[0, 1].
$$

The expected number of rejections is $c_\gamma + \psi(\alpha)$, where c_γ is the Euler-Mascheroni constant and ψ is the digamma function. Since $\psi(\alpha) \sim \log(\alpha)$ for $\alpha \to \infty$, the representation remains efficient even for fairly large values of α. To obtain a truncation bound, we use the same approach as in

Example 4.3.3:

$$\frac{B_{N,K}}{N\gamma\alpha} = \int_0^\infty F_K(\gamma\alpha\log(1+x^{-1}))(1+x)^{-\alpha-1}\mathrm{d}x$$

$$\leq \int_0^a (1+x)^{-\alpha-1}\mathrm{d}x + F_K(\gamma\alpha\log(1+a^{-1}))\int_a^\infty (1+x)^{-\alpha-1}\mathrm{d}x$$

$$\leq a + \alpha^{-1}(3\gamma\alpha\log(1+a^{-1})/K)^K(1+a)^{-\alpha}$$

$$\sim e^{-b} + \alpha^{-1}(3\gamma\alpha/K)^K b^K,$$

where $b := \log(1+a^{-1})$. Setting the two terms equal and solving for b, we conclude that

$$B_{N,K} \leq 2N\gamma\alpha e^{-KW_0(\{3\gamma\alpha\}^{-1})},$$

where W_0 is the product log function, as defined in Eq. (11).

Similarly to Example 4.2.5, for the case of $d > 0$ and $\alpha \geq 0$, we instead use $\mu(\mathrm{d}\theta) = \gamma\frac{\Gamma(\alpha+1)}{\Gamma(1-d)\Gamma(\alpha+d)}\theta^{-1-d}\mathrm{d}\theta$. Since $\mu^{\leftarrow}(u) = (\gamma' u^{-1})^{1/d}$, where $\gamma' := \gamma\frac{\Gamma(\alpha+1)}{d\Gamma(1-d)\Gamma(\alpha+d)}$, the rejection representation is

$$\Theta = \sum_{k=1}^\infty V_k \mathbb{1}(U_k \leq (1+V_k)^{-\alpha})\delta_{\psi_k}, \qquad \text{with} \qquad V_k = (\gamma'\Gamma_k^{-1})^{1/d},$$

$$U_k \overset{\text{i.i.d.}}{\sim} \mathsf{Unif}[0,1].$$

The expected number of rejections is $\frac{\gamma\alpha}{d}$, so the representation is efficient for large d, but extremely inefficient when d is small. We have

$$B_{N,K} = N\gamma\frac{\Gamma(\alpha+1)}{\Gamma(1-d)\Gamma(\alpha+d)}\int_0^\infty F_K(\gamma' x^{-d})x^{-d}(1+x)^{-1-\alpha}\mathrm{d}x.$$

Following the approach of Example 4.3.4, the integral can be upper bounded as

$$\int_0^a x^{-d}\mathrm{d}x + F_K(\gamma' a^{-d})\int_a^\infty x^{-d}(1+x)^{-1-\alpha}\mathrm{d}x$$

$$\leq (1-d)^{-1}a^{1-d} + \alpha^{-1}a^{-d}(3\gamma' K^{-1}a^{-d})^K.$$

Setting the two terms equal and solving for a, we obtain

$$B_{N,K} \leq 2N\left(\frac{\gamma\Gamma(\alpha+1)}{\Gamma(1-d)\Gamma(\alpha+d)}\right)^{\frac{(d+1)K+1}{dK+1}}(\alpha(1-d)^{dK})^{-\frac{1}{dK+1}}\left(\frac{dK}{3}\right)^{-\frac{K}{dK+1}}$$

$$\sim 2N\left(\frac{\gamma\Gamma(\alpha+1)}{\Gamma(2-d)\Gamma(\alpha+d)}\right)^{1+1/d}\left(\frac{dK}{3(1-d)^d}\right)^{-1/d} \qquad K \to \infty.$$

Decoupled Bondesson representation   It follows from Theorem 4.2.3 and the same arguments as those in the Bondesson case that if $\Theta \leftarrow \text{DB-Rep}(\gamma\alpha, \mathsf{Beta}'(1,\alpha), \xi)$, then $\Theta \sim \text{BPP}(\gamma, \alpha, 0)$. Using the trivial bound $\theta/(1+\theta) \leq \theta$ and calculations analogous to those in the beta-Bernoulli case, for $\alpha > 1$ we obtain the upper bound

$$B_{N,K} \leq 1 - \exp\left\{-N\frac{\gamma\alpha}{\alpha-1}\left(\frac{\xi}{1+\xi}\right)^K\right\} \sim N\frac{\gamma\alpha}{\alpha-1}\left(\frac{\xi}{1+\xi}\right)^K \qquad K \to \infty.$$

Power-law representation We can transform the gamma process $\Gamma\mathrm{P}(\gamma, 1, d)$ into the beta prime process $\mathrm{BPP}(\gamma, \alpha, d)$ by applying the stochastic mapping

$$\theta \mapsto u = \theta/G, \qquad G \sim \mathsf{Gam}(\alpha + d, 1), \qquad \kappa(\theta, \mathrm{d}u) = \mathsf{Gam}(\theta/u; \alpha + d, 1)\frac{\theta}{u^2}\mathrm{d}u.$$

Using Lemma 4.2.4 yields $\kappa(\Theta) \sim \mathrm{BPP}(\gamma, \alpha, d)$. Applying this result to the power-law representation $\Gamma\mathrm{P}(\gamma, 1, d)$ from Example 4.2.8 yields the novel power-law representation

$$\Theta \leftarrow \mathrm{PL\text{-}Rep}(\gamma\alpha, 1, d, \mathsf{Beta}'(1, \alpha + d)) \quad \text{implies} \quad \Theta \sim \mathrm{BPP}(\gamma, \alpha, d).$$

Using the trivial bound $\theta/(1 + \theta) \leq \theta$ and calculations analogous to those in the beta-Bernouli case, for $\alpha > 1$ we obtain the upper bound and asymptotic simplification

$$B_{N,K} \leq N\frac{\gamma\alpha}{\alpha - 1}\prod_{k=1}^{K}\frac{1 + kd}{2 + kd - d}$$

$$\sim \frac{\gamma N\alpha}{\alpha - 1}\begin{cases} 2^{-K} & d = 0 \\ \frac{\Gamma\left(\frac{2}{d}\right)}{\Gamma\left(\frac{1+d}{d}\right)}K^{1-d^{-1}} & 0 < d < 1 \end{cases} \quad K \to \infty.$$

Size-biased representation We have

$$\eta_{k1} = \eta_k = \int \pi(\theta)^{k-1}(1 - \pi(\theta))\nu(\mathrm{d}\theta) = \gamma\frac{\Gamma(\alpha + 1)\Gamma(d + k + \alpha - 1)}{\Gamma(\alpha + d)\Gamma(k + \alpha)},$$

which is the same as for the beta-Bernoulli process. Thus, the error bound is also the same as the beta-Bernoulli case.

Results and Discussion

Table 1 summarizes our truncation and simulation cost results as applied to the beta, (normalized) gamma, and beta prime processes. Results for the Bondesson representation of $\mathrm{BP}(\gamma, 1, 0)$ as well as the decoupled Bondesson representations of $\mathrm{BP}(\gamma, \lambda, 0)$ and $\Gamma\mathrm{P}(\gamma, \lambda, 0)$ were previously known, and are reproduced by our results. All other results in the table are novel to the best of the authors' knowledge. It is interesting to note that the bounds and expected costs within each of the representation classes often have the same form, aside from some constants. Across classes, however, they vary significantly, indicating that the chosen sequential representation of a process has more of an influence on the truncation error than the process itself.

Fig. 3 shows a comparison of how the truncation error bounds vary with the expected computational cost $\mathbb{E}[R]$ of simulation for the (normalized) gamma process and Poisson likelihood with $N = 5$ observations. Results shown for the thinning, rejection, and inverse-Lévy representations are computed by Monte-Carlo approximation of the formula for $B_{N,K}$ in Eq. (8), while all others use closed-form expressions from the examples in Sections 4.3 and 4.4. Note that the Bondesson and decoupled Bondesson representations do not exist when $d > 0$. Further, only those representations for which we provide closed-form bounds in the examples are shown for the normalized gamma process; we leave the numerical approximation of the results from Theorems 4.4.4 and 4.4.5 as an open problem. Similar figures for other processes (in particular, the beta-Bernoulli and beta prime-odds Bernoulli) are provided in Section 4.6. Note that all bounds presented are improved by a factor of two versus comparable past results in the literature, due to the reliance on Lemmas 4.3.1 and 4.4.1 rather than the earlier bound found in Ishwaran and James [2001].

In Fig. 3, the top row shows results for the light-tailed process ($\gamma = 1$, $\lambda = 1$, $d = 0$, and $\xi = c = \gamma\lambda$). All representations except for thinning and size-biased capture its exponential truncation error decay. This is due to the fact that the thinning representation generates increasingly many atoms of weight 0 as $K \to \infty$, and the expected number of atoms at each outer index for the size-biased representation decays as $K \to \infty$. The inverse-Lévy representation has the lowest truncation error as expected, as it is the only representation that generates a nonincreasing sequence of weights (and so must be the most efficient [Arbel and Prünster, 2017]). Based on this figure and those in Section 4.6 for other processes, it appears that the Bondesson representation typically provides the best tradeoff between simplicity and efficiency, and should be used whenever its conditions in Theorem 4.2.1 are satisfied. When the technical conditions are not satisfied, the rejection representation is a good alternative. If ease of theoretical analysis is a concern, the decoupled Bondesson representation provides comparable efficiency with the analytical simplicity of a superposition representation.

The bottom two rows of Fig. 3 show results for the heavy-tailed process ($\gamma = 1$, $\lambda = 2$, and $d \in \{0.1, 0.5\}$). The representation options are more limited, as the technical conditions
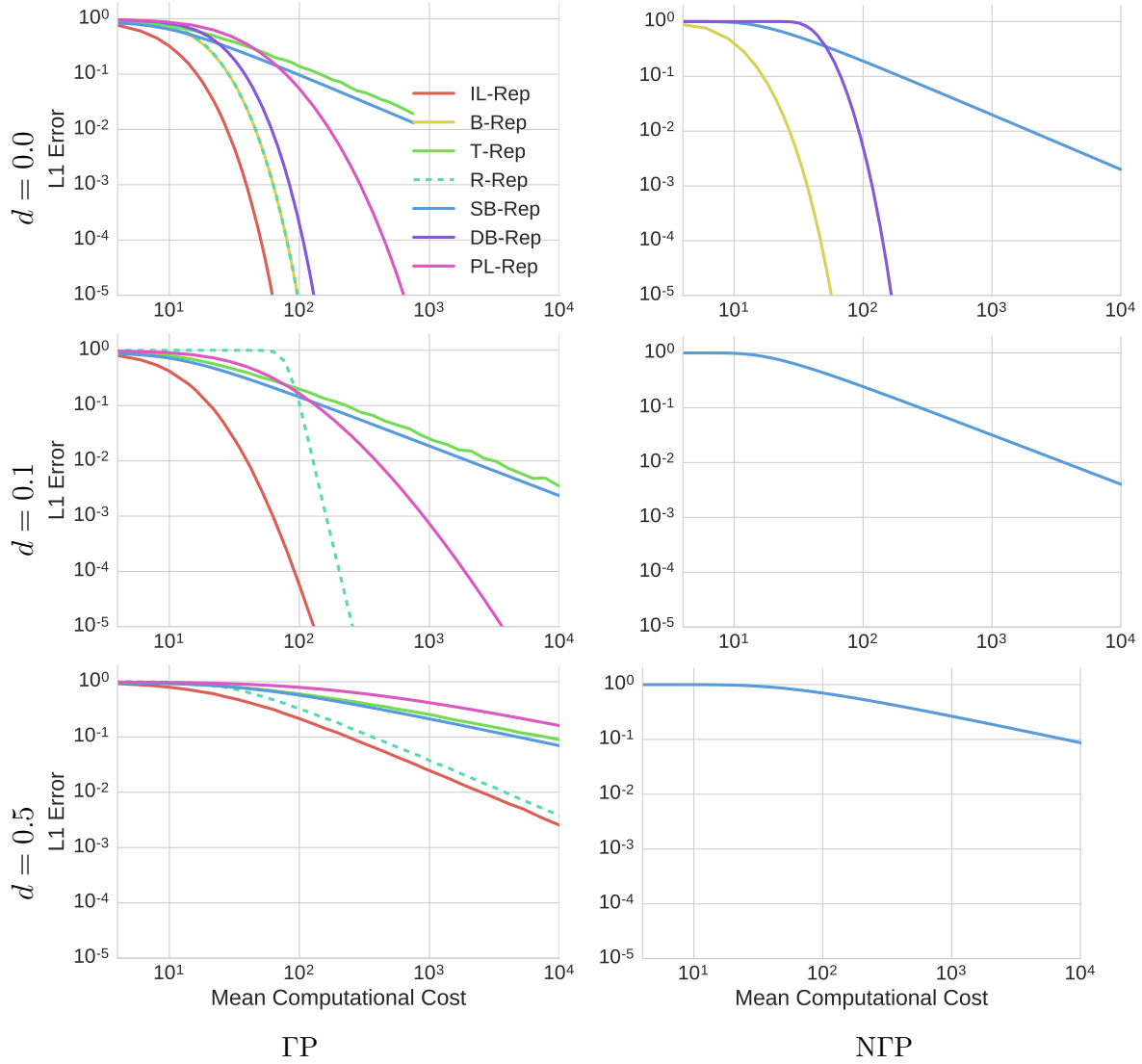
Figure 3: Truncation error bounds for representations of the (normalized) gamma-Poisson process, with $\gamma = 1$, $\lambda = 2$, and $\xi = \gamma\lambda$. The left column is for the unnormalized process, while the right column is for the normalized process. Each row displays results for a different value of the discount parameter $d \in \{0, 0.1, 0.5\}$.

Table 1: Asymptotic error bounds and simulation cost summary. Error bounds are presented up to a constant that varies between models. Be = Bernoulli, OBe = odds Bernoulli, Poi = Poisson.

Rep.	Random Measure	h	Asymptotic Error Bound	Complexity
IL	$\mathrm{LomP}(\gamma,\lambda^{-1})$	Poi	$Ne^{-KW_0\left(\{3\gamma\lambda\max(\lambda,e)\}^{-1}\right)}$	$2K$
B	$\mathrm{BP}(\gamma,\lambda\geq1,0)$ $\Gamma\mathrm{P}(\gamma,\lambda,0)$ $\mathrm{BPP}(\gamma,\lambda,0)$	Be Poi OBe	$N\gamma\left(\frac{\gamma\lambda}{\gamma\lambda+1}\right)^K$	$3K$
	$\mathrm{DP}(\gamma)$	—	$N\left(\frac{\gamma}{\gamma+1}\right)^K$	
T	—	—	See Eq. (12)	$3K$
R	$\mathrm{BP}(\gamma,\lambda,0)$ $\Gamma\mathrm{P}(\gamma,\lambda,0)$ $\mathrm{BPP}(\gamma,\lambda,0)$	Be Poi OBe	$N\begin{cases} e^{-KW_0(\{3\gamma\lambda\}^{-1})} & d=0 \\ K^{-d(1-d)} & d>0\,(\Gamma\mathrm{P}) \\ K^{-1/d} & d>0\,(\mathrm{BP},\mathrm{BPP})\end{cases}$	$3K$
DB	$\mathrm{BP}(\gamma,\lambda\geq1,0)$ $\Gamma\mathrm{P}(\gamma,\lambda,0)$ $\mathrm{BPP}(\gamma,\lambda>1,0)$	Be Poi OBe	$N\left(\frac{\xi}{\xi+1}\right)^K$	$\left(\frac{3c}{\xi}+1\right)K$
	$\mathrm{DP}(\gamma)$	—	$\frac{N\gamma}{a(\gamma-a)}\left(\frac{\xi}{\xi+a}\right)^K,\ a\in(0,1]\cap(0,\gamma)$	
SB	$\mathrm{BP}(\gamma,\lambda,d)$ $\Gamma\mathrm{P}(\gamma,\lambda,d)$ $\mathrm{BPP}(\gamma,\lambda,d)$	Be Poi OBe	NK^{d-1}	K
	$\mathrm{N}\Gamma\mathrm{P}(\gamma,\lambda,d)$	—	$N\begin{cases} K^{-1}\log K & d=0,\gamma\lambda=1 \\ K^{-\min(1,\gamma\lambda)} & d=0,\gamma\lambda\neq1 \\ K^{d-1} & d>0\end{cases}$	
PL	$\mathrm{BP}(\gamma,\lambda,d)$ $\Gamma\mathrm{P}(\gamma,\lambda,d)$ $\mathrm{BPP}(\gamma,\lambda>1,d)$	Be Poi OBe	$N\begin{cases} \left(\frac{\lambda}{\lambda+1}\right)^K & d=0\,(\mathrm{BP},\Gamma\mathrm{P}) \\ 2^{-K} & d=0\,(\mathrm{BPP}) \\ K^{1-1/d} & d>0\end{cases}$	$\frac{\gamma}{2}K^2$

of the Bondesson and decoupled Bondesson representations are not satisfied. Here the rejection representation is often the best choice due to its simplicity and competitive performance with the inverse-Lévy representation. However, one must take care to check its efficiency beforehand using Proposition 4.2.2 given a particular choice of $\mu(\mathrm{d}\theta)$. For example, the choice of $\mu(\mathrm{d}\theta)\propto\theta^{-1-d}\mathrm{d}\theta$ in the present work makes the rejection representation very inefficient when $d\ll1$ for both the gamma-Poisson (Fig. 3) and beta prime-odds Bernoulli (Fig. 2) processes, but efficient for the beta-Bernoulli process (Fig. 1). If no $\mu(\mathrm{d}\theta)$ yields reasonable results, the power-law representation is a good choice for $d\ll1$ as its truncation bound approaches the exponential decay of the light-tailed process. For larger $d>0$ the size-biased representation is a good alternative.

Based on the results in Fig. 3, it appears that there is no single dominant representation for all situations (provided the inverse-Lévy representation is intractable, as it most often is). However, as a guideline, the rejection and Bondesson representations tend to be good choices for light-tailed processes, while the rejection, size-biased, and power-law representations are good choices for heavy-tailed processes.

Conclusions

We have investigated sequential representations, truncation error bounds, and simulation algorithms for (normalized) completely random measures. In past work, the development and analysis of these tools has occurred only on an ad hoc basis. The results in the present paper, in contrast, provide a comprehensive characterization and analysis of the different types of sequential (N)CRM representations available to the practitioner. However, there are a number of remaining open questions and limitations.

First, this work does not consider the influence of observed data: all analyses assume an *a priori* perspective, as truncation is typically performed before data are incorporated via posterior inference (e.g. in variational inference for the DP mixture [Blei and Jordan, 2006] and BP latent feature model [Doshi-Velez et al., 2009]). However, analysis of *a posteriori* truncation has been studied in past work as well [Ishwaran and James, 2001, Gelfand and Kottas, 2002, Ishwaran and Zarepour, 2002]. In the language of CRMs, observations introduce a fixed-location component in the posterior process, while the unobserved traits are drawn from the (possibly normalized) ordinary component of a CRM [Ishwaran and Zarepour, 2002, Broderick et al., 2017]. We anticipate that this property makes observations reasonably simple to include: the truncation tools provided in the present paper can be used directly on the unobserved ordinary component, while the fixed-location component may be treated exactly.

In addition, there are important open questions regarding the sequential representations developed in this work. It is unknown whether generalized versions of the Bondesson and decoupled Bondesson representations can be developed for larger classes of rate measures. The power-law representation does provide a partial answer in the decoupled Bondesson case. Regarding size-biased representations, one might expect that the use of conjugate exponential family CRMs [Broderick et al., 2017] would yield a closed-form expression for the truncation bound. In all of the cases provided in this paper, this was indeed the case; the integrals were evaluated exactly and a closed-form expression was found. However, we were unable to identify a general expression applicable to all conjugate exponential family CRMs. Based on the examples provided, we conjecture that such an expression exists. Finally, fundamental connections between some of the representations were left largely unexplored in this work. This is an open area of research, although progress has been made by connecting decoupled Bondesson and size-biased representations for (hierarchies of) generalized beta processes [Roy, 2014, Sec. 6.4].

A final remark is that one of the primary uses of sequential representations in past work has been in the development of posterior inference procedures [Paisley et al., 2010, Blei and Jordan, 2006, Doshi-Velez et al., 2009]. The present work provides no guidance on which truncated representations are best paired with which inference methods. We leave this as an open direction for future research, which will require both theoretical and empirical investigation.

References

Edoardo M Airoldi, David Blei, Elena A Erosheva, and Stephen E Fienberg. *Handbook of Mixed Membership Models and Their Applications*. CRC Press, 2014.

Peter Orbanz. Conjugate projective limits. *arXiv preprint arXiv:1012.0363*, 2010.

T. Broderick, A. C. Wilson, and M. I. Jordan. Posteriors, conjugacy, and exponential families for completely random measures. *Bernoulli*, 2017.

Raffaele Argiento, Ilaria Bianchini, and Alessandra Guglielmi. A blocked gibbs sampler for ngg-mixture models via a priori truncation. *Statistics and Computing*, 26(3):641–661, 2016.

P Muliere and L Tardella. Approximating distributions of random functionals of Ferguson–Dirichlet priors. *Canadian Journal of Statistics*, 26(2):283–297, 1998.

J. F. C. Kingman. Completely random measures. *Pacific Journal of Mathematics*, 21(1):59–78, 1967.

T S Ferguson and M J Klass. A representation of independent increment processes without Gaussian components. *The Annals of Mathematical Statistics*, 43(5):1634–1643, 1972.

Nils Lid Hjort. Nonparametric bayes estimators based on beta processes in models for life history data. *The Annals of Statistics*, 18(3):1259–1294, 1990.

Yongdai Kim. Nonparametric bayesian estimators for counting processes. *The Annals of Statistics*, 27(2):562–588, 1999.

J. F. C. Kingman. Random discrete distributions. *Journal of the Royal Statistical Society B*, 37(1): 1–22, 1975.

Anders Brix. Generalized gamma measures and shot-noise cox processes. *Advances in Applied Probability*, 31:929–953, 1999.

Michalis Titsias. The infinite gamma-poisson feature model. In *Advances in Neural Information Processing Systems*, 2008.

L. F. James. Stick-breaking PG(α,ζ)-Generalized Gamma Processes. *arXiv preprint arXiv:1308.6570*, August 2013.

L Bondesson. On simulation from infinitely divisible distributions. *Advances in Applied Probability*, 14:855–869, 1982.

Jan Rosiński. On series representations of infinitely divisible random vectors. *Annals of Probability*, 18:405–430, 1990.

Jan Rosiński. Series representations of Lévy processes from the perspective of point processes. In Ole Barndorff-Nielson, Sidney Resnick, and Thomas Mikosch, editors, *Lévy processes: theory and applications*, chapter VI, pages 401–415. Bikhäuser Boston, 2001.

L. F. James. Poisson Latent Feature Calculus for Generalized Indian Buffet Processes. *arXiv preprint arXiv:1411.2936v3*, November 2014.

R. Thibaux and M I Jordan. Hierarchical beta processes and the Indian buffet process. In *International Conference on Artificial Intelligence and Statistics*, 2007.

Mingyuan Zhou, Lauren Hannah, David Dunson, and Lawrence Carin. Beta-negative binomial process and poisson factor analysis. In *Artificial Intelligence and Statistics*, 2012.

T. Broderick, L. Mackey, J. Paisley, and M. I. Jordan. Combinatorial clustering and the beta negative binomial process. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):290–306, 2015.

Mihael Perman, Jim Pitman, and Marc Yor. Size-biased sampling of poisson point processes and excursions. *Probability Theory and Related Fields*, 92(1):21–39, 1992.

Mihael Perman. Order statistics for jumps of normalised subordinators. *Stochastic Processes and their Applications*, 46(2):267–281, 1993.

L. F. James. Poisson Process Partition Calculus with applications to Exchangeable models and Bayesian Nonparametrics. *arXiv preprint arXiv:0205093*, May 2002.

J. Pitman. Poisson-Kingman partitions. *Lecture Notes-Monograph Series*, 2003.

Eugenio Regazzini, Antonio Lijoi, and Igor Prünster. Distributional results for means of normalized random measures with independent increments. *The Annals of Statistics*, 31(2):560–585, 2003.

Antonio Lijoi and Igor Prünster. Models beyond the dirichlet process. In Nils Lid Hjort, Chris Holmes, Peter Müller, and Stephen Walker, editors, *Bayesian Nonparametrics*, pages 80–136. Cambridge University Press, 2010.

L. F. James, Antonio Lijoi, and Igor Prünster. Posterior Analysis for Normalized Random Measures with Independent Increments. *Scandinavian Journal of Statistics*, 36(1):76–97, 2009.

Thomas S Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 1973.

Jayaram Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639–650, 1994.

H. Ishwaran and L. F. James. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453):161–173, 2001.

H. Ishwaran and M Zarepour. Exact and approximate sum representations for the Dirichlet process. *Canadian Journal of Statistics*, 30(2):269–283, 2002.

D. M. Blei and M. I. Jordan. Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, 1(1):121–144, 2006.

J W Paisley, D. M. Blei, and M I Jordan. Stick-breaking beta processes and the Poisson process. *International Conference on Artificial Intelligence and Statistics*, 2012.

F. Doshi-Velez, K. T. Miller, J. Van Gael, and Y. W. Teh. Variational inference for the Indian buffet process. In *International Conference on Artificial Intelligence and Statistics*, 2009.

Daniel Roy. The continuum-of-urns scheme, generalized beta and Indian buffer processes, and hierarchies thereof. *arXiv preprint arXiv:1501.00208*, 2014.

Anirban Roychowdhury and Brian Kulis. Gamma Processes, Stick-Breaking, and Variational Inference. In *International Conference on Artificial Intelligence and Statistics*, 2015.

Antonio Lijoi, Ramsés Mena, and Igor Prünster. Bayesian nonparametric analysis for a generalized dirichlet process prior. *Statistical Inference for Stochastic Processes*, 8:283–309, 2005.

H. Ishwaran and L. F. James. Approximate Dirichlet Process Computing in Finite Normal Mixtures: Smoothing and Prior Information. *Journal of Computational and Graphical Statistics*, 11 (3):508–532, September 2002.

Emil Julius Gumbel. *Statistical theory of extreme values and some practical applications: A series of lectures*. National Bureau of Standards Applied Mathematics Series 33. U.S. Government Printing Office, Washington, D.C., 1954.

Chris Maddison, Daniel Tarlow, and T. P. Minka. A* Sampling. In *Advances in Neural Information Processing Systems*, 2014.

Y W Teh and D. Görür. Indian buffet processes with power-law behavior. In *Advances in Neural Information Processing Systems*, 2009.

T. Broderick, M. I. Jordan, and J. Pitman. Beta processes, stick-breaking and power laws. *Bayesian Analysis*, 7(2):439–476, 2012.

Milton Abramowitz and Irene Stegun, editors. *Handbook of Mathematical Functions*. Dover Publications, 1964.

D Banjevic, H. Ishwaran, and M Zarepour. A recursive method for functionals of Poisson processes. *Bernoulli*, 8(3):295–311, 2002.

J W Paisley, A K Zaas, C W Woods, G S Ginsburg, and Lawrence Carin. A stick-breaking construction of the beta process. *International Conference on Machine Learning*, 2010.

Julyan Arbel and Igor Prünster. A moment-matching Ferguson & Klass algorithm. *Statistics and Computing*, 27(1):3–17, 2017.

J. F. C. Kingman. *Poisson Processes*. Oxford Studies in Probability. Oxford University Press, 1993.

Antonio Lijoi and Igor Prünster. On a normalized random measure with independent increments relevant to Bayesian nonparametric inference. In *Proceedings of the 13th European Young Statisticians Meeting*, pages 123–124. Bernoulli Society, 2003.

Antonio Lijoi, Ramsés Mena, and Igor Prünster. Controlling the reinforcement in Bayesian nonparametric mixture models. *Journal of the Royal Statistical Society B*, 69(4):715–740, 2007.

Y W Teh, D. Görür, and Z. Ghahramani. Stick-breaking construction for the Indian buffet process. In *International Conference on Artificial Intelligence and Statistics*, 2007.

Alan Gelfand and Athanasios Kottas. A computational approach for nonparametric bayesian inference under dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 11(2):289–305, 2002.

Walter Gautschi. Some elementary inequalities relating to the gamma and incomplete gamma function. *Journal of Mathematics and Physics*, 38(1):77–81, 1959.

Technical Appendices

A.1 Technical lemmas

Lemma A.1.1. *If Y_k is a uniformly bounded, non-negative sequence of random variables such that $\lim_{k \to \infty} \mathbb{E}[Y_k] = 0$, then $Y_k \xrightarrow{P} 0$.*

Proof. Without loss of generality we assume that $Y_k \in [0, 1]$ a.s. Then for all $\epsilon, \delta > 0$, by hypothesis there exists k' such that for all $k \geq k'$, $\mathbb{E}[Y_k] \leq \epsilon\delta$. It then follows from Markov's inequality that for all $k \geq k'$, $\mathbb{P}(Y_k > \epsilon) \leq \delta$. $\square$

Lemma A.1.2. *If μ is a non-atomic measure on $\mathbb{R}^d$, then for any $x \in \mathbb{R}^d$ and $\delta > 0$, there exists $\epsilon_{x,\delta} > 0$ such that $\mu(\{y \in \mathbb{R}^d \mid \|x - y\|_2 \leq \epsilon_{x,\delta}\}) \leq \delta$.*

Proof. Without loss of generality let $x = 0$. Suppose the implication does not hold. Then there exists $\delta > 0$ such that for all $\epsilon > 0$, $\mu(B_\epsilon) > \delta$, where $B_\epsilon := \{y \in \mathbb{R}^d \mid \|y\|_2 \leq \epsilon\}$. Let ϵ_n be a sequence such that $\epsilon_n \to 0$ as $n \to \infty$. Then by continuity $\mu(\{0\}) = \lim_{n \to \infty} \mu(B_{\epsilon_n}) > \delta$, hence μ is a atomic, which is a contradiction. $\square$

Lemma A.1.3. *If $\nu(\mathrm{d}\theta)$, an absolutely continuous σ-finite measure on $\mathbb{R}_+$, and continuous $\phi : \mathbb{R}_+ \to [0, 1]$ satisfy*

$$\nu(\mathbb{R}_+) = \infty, \qquad \int \min(1, \theta)\nu(\mathrm{d}\theta) < \infty, \qquad and \qquad \int \phi(\theta)\nu(\mathrm{d}\theta) < \infty,$$

then

$$\lim_{\theta \to 0} \phi(\theta) = \phi(0) = 0.$$

Proof. $\int \min(1, \theta)\nu(\mathrm{d}\theta) < \infty$ implies that $\int_0^1 \theta\nu(\mathrm{d}\theta) < \infty$ and $\int_1^\infty \nu(\mathrm{d}\theta) < \infty$. But $\nu(\mathbb{R}_+) = \infty$, so $\int_0^1 \nu(\mathrm{d}\theta) = \infty$. In fact, for all $\epsilon > 0$, $\nu([0, \epsilon]) = \infty$ since otherwise $\int_\epsilon^1 \nu(\mathrm{d}\theta) = \infty \implies \int_\epsilon^1 \theta\nu(\mathrm{d}\theta) = \infty$, a contradiction. Since ϕ is continuous and has bounded range, $\lim_{\theta \to 0} \phi(\theta) = c$ exists and is finite. Assume $c > 0$, so $\exists \epsilon > 0$ such that $\forall \epsilon' < \epsilon$, $|\phi(\epsilon') - c| < c/2$, and in particular $\phi(\epsilon') > c/2$. Thus, for any $\theta' < \epsilon$, $\int_0^{\theta'} \phi(\theta)\nu(\mathrm{d}\theta) \geq c/2 \int_0^{\theta'} \nu(\mathrm{d}\theta) = \infty$, a contradiction. Thus, $c = 0$. $\square$

Lemma A.1.4. *Assume $\phi(x)$ is a twice continuously differentiable function with the following properties:*

1. $\phi''(x)/\phi'(x) \to 0$ *as* $x \to \infty$

2. *for all* $\delta > 0$ *there exists* $B_\delta > 0$ *such that for any increasing sequence* $(x_n)_{n=1}^\infty$,

$$\limsup_{n\to\infty} \sup_{y\in[x_n,x_n+\delta]} \frac{\phi''(y)}{\phi''(x_n)} = B_\delta < \infty.$$

Then for any constant $c > 0$ *and any increasing sequence* $(x_n)_{n=1}^\infty$,

$$\phi(x_n + c) - \phi(x_n) \sim c\phi'(x_n) \qquad\qquad \textit{for } n \to \infty.$$

Proof. A second-order Taylor expansion of $\phi(x_n + c)$ about x_n yields

$$\phi(x_n + c) - \phi(x_n) = c\phi'(x_n) + \frac{c^2}{2}\phi''(x_n^*),$$

where $x_n^* \in [x_n, x_n + c]$. Our assumptions on ϕ ensure that,

$$\lim_{n\to\infty} \frac{\phi''(x_n^*)}{\phi'(x_n)} \le \lim_{n\to\infty} \frac{B_c\phi''(x_n)}{\phi'(x_n)} = 0$$

and hence

$$\lim_{n\to\infty} \frac{\phi(x_n + c) - \phi(x_n)}{c\phi'(x_n)} = \lim_{n\to\infty} 1 + \frac{c}{2}\frac{\phi''(x_n^*)}{\phi'(x_n)} = 1.$$

$\square$

Lemma A.1.5 (Gautschi [1959])**.**

$$(1 + x)^{d-1} \le \frac{\Gamma(x + d)}{\Gamma(x + 1)} \le x^{d-1} \qquad 0 \le d \le 1, x \ge 1,$$

and thus for $0 \le d \le 1$,

$$\frac{\Gamma(x + d)}{\Gamma(x + 1)} \sim x^{d-1} \qquad\qquad x \to \infty.$$

Lemma A.1.6. *For* $\alpha > 0$ *and* $x \ge -1$,

$$\sum_{m=1}^{M} \frac{\Gamma(\alpha + m + x)}{\Gamma(\alpha + m)} = \begin{cases} \frac{1}{1+x}\left(\frac{\Gamma(\alpha+M+x+1)}{\Gamma(\alpha+M)} - \frac{\Gamma(\alpha+x+1)}{\Gamma(\alpha)}\right) & x > -1 \\ \psi(\alpha + M) - \psi(\alpha) & x = -1 \end{cases}$$

where $\psi(\cdot)$ *is the digamma function.*

Proof. When $M = 1$ and $x > -1$, analyzing the right hand side yields

$$\frac{\Gamma(\alpha + M + x + 1)}{\Gamma(\alpha + M)} - \frac{\Gamma(\alpha + x + 1)}{\Gamma(\alpha)} = \frac{\Gamma(\alpha + x + 2)}{\Gamma(\alpha + 1)} - \frac{\Gamma(\alpha + x + 1)}{\Gamma(\alpha)}$$

$$= \frac{\Gamma(\alpha + x + 1)}{\Gamma(\alpha)}\left(\frac{\alpha + x + 1}{\alpha} - 1\right)$$

$$= (x + 1)\frac{\Gamma(\alpha + x + 1)}{\Gamma(\alpha + 1)}.$$

43

By induction, supposing that the result is true for $M - 1 \geq 1$ and $x > -1$,

$$\sum_{m=1}^{M} \frac{\Gamma(\alpha + m + x)}{\Gamma(\alpha + m)} = \sum_{m=1}^{M-1} \frac{\Gamma(\alpha + m + x)}{\Gamma(\alpha + m)} + \frac{\Gamma(\alpha + M + x)}{\Gamma(\alpha + M)}$$

$$= \frac{1}{1 + x} \left(\frac{\Gamma(\alpha + M + x)}{\Gamma(\alpha + M - 1)} - \frac{\Gamma(\alpha + x + 1)}{\Gamma(\alpha)} \right) + \frac{\Gamma(\alpha + M + x)}{\Gamma(\alpha + M)}$$

$$= \frac{\Gamma(\alpha + M + x)}{\Gamma(\alpha + M - 1)} \frac{\alpha + M + x}{(1 + x)(\alpha + M - 1)} - \frac{\Gamma(\alpha + x + 1)}{(1 + x)\Gamma(\alpha)}$$

$$= \frac{1}{1 + x} \left(\frac{\Gamma(\alpha + M + x + 1)}{\Gamma(\alpha + M)} - \frac{\Gamma(\alpha + x + 1)}{\Gamma(\alpha)} \right).$$

This demonstrates the desired result for $x > -1$. Next, when $x = -1$, we have that

$$\sum_{m=1}^{M} \frac{\Gamma(\alpha + m - 1)}{\Gamma(\alpha + m)} = \sum_{m=1}^{M} \frac{1}{\alpha + m - 1}.$$

We proceed by induction once again. For $M = 1$, using the recurrence relation $\psi(x + 1) = \psi(x) + x^{-1}$ [Abramowitz and Stegun, 1964, Chapter 6], the right hand side evaluates to

$$\psi(\alpha + 1) - \psi(\alpha) = \psi(\alpha) + \alpha^{-1} - \psi(\alpha) = \alpha^{-1}.$$

Supposing that the result is true for $M - 1 \geq 1$ and $x = -1$,

$$\sum_{m=1}^{M} \frac{1}{\alpha + m - 1} = \sum_{m=1}^{M-1} \frac{1}{\alpha + m - 1} + \frac{1}{\alpha + M - 1}$$

$$= \psi(\alpha + M - 1) - \psi(\alpha) + \frac{1}{\alpha + M - 1}$$

$$= \psi(\alpha + M) - \psi(\alpha),$$

demonstrating the result for $x = -1$. $\qquad\square$

Lemma A.1.7. *For $a > 0$, $d \in \mathbb{R}$, and $x_n \to \infty$,*

$$\frac{\mathrm{d}}{\mathrm{d}x_n} \frac{\Gamma(a + x_n + d)}{\Gamma(a + x_n)} \sim dx_n^{d-1}.$$

Proof. We have

$$\frac{\mathrm{d}}{\mathrm{d}x} \frac{\Gamma(a + x + d)}{\Gamma(a + x)} = \frac{\Gamma(a + x + d)}{\Gamma(a + x)} (\psi(a + x + d) - \psi(a + x)),$$

where ψ is the digamma function. Using Lemma A.1.4 and the asymptotic expansion of ψ' [Abramowitz and Stegun, 1964, Chapter 6], we obtain

$$\psi(a + x_n + d) - \psi(a + x_n) \sim d\psi'(a + x_n) \sim \frac{d}{x_n + a} \sim dx_n^{-1}.$$

Since

$$\frac{\Gamma(a + x_n + d)}{\Gamma(a + x_n)} \sim (a + x_n)^d \sim x_n^d,$$

using Lemma A.1.9(2) with the previous two displays yields the result. $\qquad\square$

Lemma A.1.8. *For $0 \leq d < 1$,*

$$\int_0^\infty \left(e^{-t\theta} - 1\right) \theta^{-1-d} e^{-\lambda\theta} \mathrm{d}\theta = \begin{cases} \Gamma(-d)\left((\lambda+t)^d - \lambda^d\right) & 0 < d < 1 \\ \log\left(\frac{\lambda}{t+\lambda}\right) & d = 0 \end{cases} .$$

Proof. By integration by parts and the standard gamma integral,

$$\int_0^\infty \left(e^{-t\theta} - 1\right) \theta^{-1-d} e^{-\lambda\theta} \mathrm{d}\theta = \int_0^\infty \left[(\lambda+t)e^{-(\lambda+t)\theta} - \lambda e^{-\lambda\theta}\right] \frac{\theta^{-d}}{-d} \mathrm{d}\theta$$
$$= \Gamma(-d)\left((\lambda+t)^d - \lambda^d\right).$$

Taking the limit as $d \to 0$ via L'Hospital's rule yields

$$\lim_{d \to 0} \Gamma(-d)\left((\lambda+t)^d - \lambda^d\right) = \log\left(\frac{\lambda}{\lambda+t}\right).$$

$\square$

Lemma A.1.9 (Standard asymptotic equivalence properties)**.**

1. *If $a_n \sim b_n$ and $b_n \sim c_n$, then $a_n \sim c_n$.*

2. *If $a_n \sim b_n$ and $c_n \sim d_n$, then $a_n c_n \sim b_n d_n$.*

3. *If $a_n \sim b_n$, $c_n \sim d_n$ and $a_n c_n > 0$, then $a_n + c_n \sim b_n + d_n$.*

A.2 Proofs of sequential representation results

A.2.1 Correctness of B-Rep, DB-Rep, and PL-Rep

Proof of Theorem 4.2.1. First, we show that $g_\nu(v)$ is a density. Since $v\nu(v)$ is nondecreasing, $\frac{\mathrm{d}}{\mathrm{d}v}[v\nu(v)]$ exists almost everywhere, $\frac{\mathrm{d}}{\mathrm{d}v}[v\nu(v)] \leq 0$ and hence $g_\nu(v) \geq 0$. Furthermore,

$$\int_0^\infty g_\nu(v)\mathrm{d}v = -c_\nu^{-1} \int_0^\infty \frac{\mathrm{d}}{\mathrm{d}v}[v\nu(v)]\mathrm{d}v = -c_\nu^{-1} v\nu(v)\Big|_{v=0}^\infty = 1,$$

where the final equality follows from the assumed behavior of $v\nu(v)$ at 0 and ∞. Since for a partition $A_1, \ldots, A_n$, the random variables $\Theta(A_1), \ldots, \Theta(A_n)$ are independent, it suffices to show that for any measurable set A (with complement $\bar{A}$), the random variable $\Theta(A)$ has the correct characteristic function. Define the family of random measures

$$\Theta_t = \sum_{k=1}^\infty V_k e^{-(\Gamma_k+t)/c_\nu} \delta_{\psi_k}, \qquad\qquad t \geq 0,$$

so $\Theta_0 = \Theta$. Conditioning on Γ_{11},

$$(\Theta_t(A) \mid \Gamma_1 = u) \overset{\mathcal{D}}{=} V_1 e^{-(u+t)/c_\nu} \mathbb{1}[\psi_1 \in A] + \Theta_{t+u}(A),$$

and note that the two terms on the left hand side are independent. We can thus write the characteristic function of $\Theta_t(A)$ as

$$
\begin{aligned}
\varphi(\xi, t, A) &:= \mathbb{E}[e^{i\xi\Theta_t(A)}] \\
&= \mathbb{E}[\mathbb{E}[e^{i\xi\Theta_t(A)} \mid \Gamma_1 = u]] \\
&= \mathbb{E}[\mathbb{E}[e^{i\xi V_1 e^{-(u+t)/c_\nu} \mathbb{1}[\psi_1 \in A]} e^{i\xi\Theta_{t+u}(A)} \mid \Gamma_1 = u]] \\
&= \mathbb{E}[\mathbb{E}[(G(A)e^{i\xi V_1 e^{-(u+t)/c_\nu}} + G(\bar{A}))\varphi(\xi, t+u, A) \mid \Gamma_1 = u]] \\
&= G(A) \int_0^\infty \int_0^\infty e^{i\xi v e^{-(u+t)/c_\nu}} \varphi(\xi, t+u, A)g_\nu(v)e^{-u}\, du\, dv \\
&\quad + G(\bar{A}) \int_0^\infty \varphi(\xi, t+u, A)e^{-u}\, du,
\end{aligned}
$$

where $\bar{A}$ is the complement of A. Multiplying both sides by e^{-t} and making the change of variable $w = u + t$ yields

$$
\begin{aligned}
e^{-t}\varphi(\xi, t, A) &= G(A) \int_t^\infty \int_0^\infty e^{i\xi v e^{-w/c_\nu}} \varphi(\xi, w, A)g_\nu(v)e^{-w}\, dv\, dw \\
&\quad + G(\bar{A}) \int_t^\infty \varphi(\xi, w, A)e^{-w}\, dw \\
&= G(A) \int_t^\infty \varphi_{g_\nu}(\xi e^{-w/c_\nu})\varphi(\xi, w, A)e^{-w} dw \\
&\quad + G(\bar{A}) \int_t^\infty \varphi(\xi, w, A)e^{-w} dw,
\end{aligned}
$$

where $\varphi_{g_\nu}(a) := \int_0^\infty e^{iav}g_\nu(v)\, dv$ is the characteristic function of a random variable with density g_ν. Differentiating both sides with respect to t and rearranging yields

$$
\begin{aligned}
\frac{\partial\varphi(\xi, t, A)}{\partial t} &= \varphi(\xi, t, A) - G(A)\varphi_{g_\nu}(\xi e^{-t/c_\nu})\varphi(\xi, t, A) - (1 - G(A))\varphi(\xi, t, A) \\
&= \varphi(\xi, t, A)G(A)(1 - \varphi_{g_\nu}(\xi e^{-t/c_\nu})),
\end{aligned}
$$

so we conclude that

$$
\varphi(\xi, t, A) = \exp\left(-G(A) \int_t^\infty (1 - \varphi_{g_\nu}(\xi e^{-u/c_\nu}))\, du\right).
$$

Using integration by parts and the definition of g_ν, rewrite

$$
\begin{aligned}
\varphi_{g_\nu}(a) &= -c_\nu^{-1} \int_0^\infty \frac{d}{dv}[v\nu(v)]e^{iav}\, dv \\
&= -c_\nu^{-1} v\nu(v)e^{iav}\Big|_{v=0}^\infty + c_\nu^{-1} \int_0^\infty iav\nu(v)e^{iav}\, dv \\
&= 1 + \int_0^\infty \frac{iav}{c_\nu}\nu(v)e^{iav}\, dv,
\end{aligned}
$$

where the final equality follows from the assumed behavior of $v\nu(v)$ at 0 and ∞. Combining the previous two displays and setting $t = 0$ concludes the proof:

$$\varphi(\xi, 0, A) = \exp\left(-G(A)\int_0^\infty \int_0^\infty \frac{i\xi v}{c_\nu} e^{-u/c_\nu} e^{i\xi v e^{-u/c_\nu}} \nu(v)\,\mathrm{d}v\,\mathrm{d}u\right)$$

$$= \exp\left(-G(A)\int_0^\infty \int_0^\infty \frac{\partial}{\partial u}\left[-e^{i\xi v e^{-u/c_\nu}}\right]\nu(v)\,\mathrm{d}u\,\mathrm{d}v\right)$$

$$= \exp\left(-G(A)\int_0^\infty (e^{i\xi v} - 1)\nu(v)\,\mathrm{d}v\right).$$

$\square$

Proof of Theorem 4.2.3. It was already shown that $g_\nu(v)$ is a density. Let $\Theta_k' = \sum_{i=1}^{C_k} \theta_{ki}\delta_{\psi_{ki}}$, so $\Theta = \sum_{k=1}^\infty \Theta_k'$. Each Θ_k' is a CRM with rate measure $\frac{c_\nu}{\xi}\nu_k'(\mathrm{d}\theta)$, where $\nu_k'(\mathrm{d}\theta)$ is the law of θ_{ki}. Using the product distribution formula we have

$$\nu_k'(\mathrm{d}\theta) = \int_0^1 \frac{\xi^k}{\Gamma(k)}(-\log w)^{k-1}w^{\xi-2}g_\nu(\theta/w)\,\mathrm{d}w\,\mathrm{d}\theta. \tag{27}$$

Let $G_\nu(v) = \int_0^v g_\nu(x)\,\mathrm{d}x$ be the cdf derived from g_ν. From the preceding arguments, conclude that the rate measure of Θ is

$$\nu'(\mathrm{d}\theta) := \frac{c_\nu}{\xi}\sum_{k=1}^\infty \nu_k'(\mathrm{d}\theta)$$

$$= c_\nu\int_0^1 \sum_{k=1}^\infty \frac{\xi^{k-1}}{\Gamma(k)}(-\log w)^{k-1}w^{\xi-2}g_\nu(\theta/w)\,\mathrm{d}w\,\mathrm{d}\theta$$

$$= c_\nu\int_0^1 \xi w^{-2}g_\nu(\theta/w)\,\mathrm{d}w\,\mathrm{d}\theta$$

$$= c_\nu\int_0^1 \xi\theta^{-1}\frac{\partial}{\partial w}[-G_\nu(\theta/w)]\,\mathrm{d}w\,\mathrm{d}\theta$$

$$= -c_\nu\theta^{-1}G_\nu(\theta/w)\Big|_{w=0}^1\,\mathrm{d}\theta$$

$$= c_\nu\theta^{-1}(1 - G_\nu(\theta))\,\mathrm{d}\theta.$$

The cdf can be rewritten as

$$1 - G_\nu(\theta) = 1 + c_\nu^{-1}\int_0^\theta \frac{\mathrm{d}}{\mathrm{d}x}[x\nu(x)]\,\mathrm{d}x = 1 + c_\nu^{-1}x\nu(x)\Big|_{x=0}^\theta = c_\nu^{-1}\theta\nu(\theta).$$

Combining the previous two displays, conclude that $\nu'(\mathrm{d}\theta) = \nu(\theta)\,\mathrm{d}\theta$. $\square$

Proof of Theorem 4.2.5. Since the power-law representation in Eq. (5) for the case when $V_{ki} = 1$ almost surely was previously shown to be $\mathrm{BP}(\gamma, \alpha, d)$ [Broderick et al., 2012], we simply apply the stochastic mapping result in Lemma 4.2.4 with $\kappa(\theta, \mathrm{d}u) = \theta^{-1}g_\nu(u\theta^{-1})\mathrm{d}u$, where $\theta^{-1}g_\nu(u\theta^{-1})$ is the density of $U = V\theta \mid \theta$ under $V \sim g_\nu$. $\square$

A.2.2 Proof of the expected number rejections of the R-Rep

Proof of Proposition 4.2.2. We have

$$
\begin{aligned}
\mathbb{E}\left[\sum_{k=1}^{\infty} \mathbb{1}(\theta_k = 0)\right] &= \mathbb{E}\left[\sum_{k=1}^{\infty} \mathbb{1}\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(V_k) \geq U_k\right)\right] \\
&= \mathbb{E}\left[\sum_{k=1}^{\infty}\left(1 - \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(V_k)\right)\right] \\
&= \int\left(1 - \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)\right)\mu(\mathrm{d}x),
\end{aligned}
$$

where the equalities follow from the definition of θ_k, integrating out U_k, and applying Campbell's theorem. □

A.2.3 Power-law behavior of the PL-Rep

We now formalize the sense in which power-law representations do in fact produce power-law behavior. Let $Z_n \,|\, \Theta \overset{\text{i.i.d.}}{\sim} \mathrm{LP}(\mathsf{Poiss}, \Theta)$ and $y_k := \sum_{n=1}^{N} \mathbb{1}[z_{nk} \geq 1]$. We analyze the number of non-zero features after N observations,

$$
K_N := \sum_{k=1}^{\infty} \mathbb{1}[y_k \geq 1],
$$

and the number of features appearing $j > 1$ times after N observations,

$$
K_{N,j} := \sum_{k=1}^{\infty} \mathbb{1}[y_k = j].
$$

In their power law analysis of the beta process, Broderick et al. [2012] use a Bernoulli likelihood process. However, the Bernoulli process is only applicable if $\theta_k \in [0, 1]$, whereas in general $\theta_k \in \mathbb{R}_+$. Replacing the Bernoulli process with a Poisson likelihood process is a natural choice since $\mathbb{1}[z_{nk} \geq 1] \sim \mathsf{Bern}(1 - e^{-\theta_k})$, and asymptotically $1 - e^{-\theta_k} \sim \theta_k$ a.s. for $k \to \infty$ since $\lim_{k \to \infty} \theta_k = 0$ a.s. Thus, the Bernoulli and Poisson likelihood processes behave the same asymptotically, which is what is relevant to our asymptotic analysis. We are therefore able to show that all CRMs with power-law representations, not just the beta process, have what Broderick et al. [2012] call Types I and II power law behavior. Our only condition is that the tails of g are not too heavy.

Theorem A.2.1. *Assume that g is a continuous density such that for some $\epsilon > 0$,*

$$
g(x) = O(x^{-1-d-\epsilon}). \tag{28}
$$

Then for $\Theta \leftarrow \mathrm{PL\text{-}Rep}(\gamma, \alpha, d, g)$ with $d > 0$, there exists a constant C depending on γ, α, d, and g such that, almost surely,

$$
\begin{aligned}
K_N &\sim \Gamma(1 - d)CN^d, & N \to \infty \\
K_{N,j} &\sim \frac{d\,\Gamma(j - d)}{j!}CN^d, & N \to \infty \qquad (j > 1).
\end{aligned}
$$

In order to prove Theorem A.2.1, we require a number of additional definitions and lemmas. Our approach follows that in Broderick et al. [2012], which the reader is encouraged to consult for more details and further discussion of power law behavior of CRMs. Throughout this section, $\Theta \leftarrow \text{PL-Rep}(\gamma, \alpha, d, g)$ with $d > 0$. By Lemma 4.2.4, $\Theta \sim \text{CRM}(\nu)$, where

$$\nu(\mathrm{d}\theta) := \int g(\theta/u)u^{-1}\nu_{\text{BP}}(\mathrm{d}u)\,\mathrm{d}\theta$$

and $\nu_{\text{BP}}(\mathrm{d}\theta)$ is the rate measure for $\text{BP}(\gamma, \alpha, d)$. Let Π_k be a homogeneous Poisson point process on $\mathbb{R}_+$ with rate θ_k and define

$$K(t) := \sum_{k=1}^{\infty} \mathbb{1}[|\Pi_k \cap [0, t]| > 0]$$

$$K_j(t) := \sum_{k=1}^{\infty} \mathbb{1}[|\Pi_k \cap [0, t]| = j].$$

Furthermore, for $N \in \mathbb{N}$, let

$$\Phi_N := \mathbb{E}[K_N] \qquad \text{and} \qquad \Phi_{N,j} := \mathbb{E}[K_{N,j}] \quad (j > 1)$$

and for $t > 0$, let

$$\Phi(t) := \mathbb{E}[K(t)] \qquad \text{and} \qquad \Phi_j(t) := \mathbb{E}[K_j(t)] \quad (j > 1).$$

If follows from Campbell's Theorem [Kingman, 1993] that

$$\Phi(t) = \mathbb{E}\left[\sum_k (1 - e^{-t\theta_k})\right] = \int (1 - e^{-t\theta})\nu(\mathrm{d}\theta)$$

$$\Phi_N = \mathbb{E}\left[\sum_k (1 - e^{-N\theta_k})\right] = \Phi(N)$$

$$\Phi_j(t) = \binom{N}{j}\mathbb{E}\left[\sum_k \frac{t^j(1 - e^{-\theta_k})^j}{j!}e^{-t\theta_k}\right] = \frac{t^j}{j!}\int (1 - e^{-\theta})^j e^{-t\theta}\nu(\mathrm{d}\theta)$$

$$\Phi_{N,j} = \binom{N}{j}\mathbb{E}\left[\sum_k (1 - e^{-\theta_k})^j e^{-(N-j)\theta_k}\right] = \binom{N}{j}\int (1 - e^{-\theta})^j e^{-(N-j)\theta}\nu(\mathrm{d}\theta).$$

The first lemma characterizes the power law behavior of $\Phi(t)$ and $\Phi_j(t)$. A *slowly varying function* ℓ satisfies $\ell(ax)/\ell(x) \to 1$ as $x \to \infty$ for all $a > 0$.

Lemma A.2.2 (Broderick et al. [2012], Proposition 6.1)**.** *If for some $d \in (0, 1)$, $C > 0$, and slowly varying function ℓ,*

$$\bar{\nu}[0, x] := \int_0^x \theta\nu(\mathrm{d}\theta) \sim \frac{d}{1-d}C\ell(1/x)x^{1-d}, \qquad x \to 0, \qquad (29)$$

then

$$\Phi(t) \sim \Gamma(1 - d)Ct^d, \qquad\qquad t \to \infty$$

$$\Phi_j(t) \sim \frac{d\,\Gamma(j - d)}{j!}Ct^d, \qquad\qquad t \to \infty \qquad (j > 1).$$

49

Transferring the power law behavior from $\Phi(t)$ to Φ_N is trivial since $\Phi(N) = \Phi_N$. The next lemma justifies transferring the power law behavior from $\Phi_j(t)$ to $\Phi_{N,j}$.

Lemma A.2.3 (Broderick et al. [2012], Lemmas 6.2 and 6.3). *If ν satisfies Eq. (1), then*

$$K(t) \uparrow \infty \, a.s., \qquad \Phi(t) \uparrow \infty, \qquad \Phi(t)/t \downarrow 0.$$

Furthermore,

$$|\Phi_{N,j} - \Phi_j(N)| < \frac{C_j}{N} \max\{\Phi_j(N), \Phi_{j+2}(N)\} \to 0.$$

The final lemma confirms that the asymptotic behaviors of K_N and $K_{N,j}$ is almost surely the same as the expectations of K_N and $K_{N,j}$.

Lemma A.2.4. *Assume ν satisfies Eq. (1) and that for some $d \in (0,1)$, $C > 0$, $C_j > 0$, and slowly varying functions ℓ, ℓ', $\Phi(t) \sim C\ell(t)t^d$ and $\Phi_j(t) \sim C_j\ell(t)t^d$. Then for $N \to \infty$, almost surely*

$$K_n \sim \Phi_N \qquad and \qquad \sum_{i<j} K_{N,i} \sim \sum_{i<j} \Phi_{N,i}.$$

Proof of Theorem A.2.1. Combining the three lemmas, the result follows as soon as we show that $\nu(\mathrm{d}\theta)$ satisfies Eq. (29). C will be a constant that may change from line to line. We begin by rewriting $\nu(\mathrm{d}\theta)$ using the change of variable $w = \theta(u^{-1} - 1)$:

$$\nu(\mathrm{d}\theta) = C \int_0^1 g(\theta/u)u^{-2-d}(1-u)^{\alpha+d-1} \, \mathrm{d}u \, \mathrm{d}\theta$$

$$= C\theta^{-1-d} \int_0^\infty g(w+\theta)\frac{w^{\alpha+d-1}}{(w+\theta)^{\alpha-1}} \, \mathrm{d}w \, \mathrm{d}\theta.$$

Since $g(x)$ is integrable and continuous, for $x \in [0,1]$, it is upper-bounded by the non-integrable function $C_0 x^{-1}$ for some $C_0 > 0$. Combining this upper bound with Eq. (28) yields $g(x) \leq \phi(x) := C_0 x^{-1}\mathbb{1}[x \leq 1] + C_1 x^{-1-d-\epsilon}\mathbb{1}[x > 1]$ for some $C_1 > 0$, so

$$g(w+\theta)\frac{w^{\alpha+d-1}}{(w+\theta)^{\alpha-1}} \leq \phi(w)w^d.$$

Since $\phi(w)w^d$ is integrable, by dominated convergence the limit

$$L = \lim_{\theta \to 0} \int_0^\infty g(w+\theta)\frac{w^{\alpha+d-1}}{(w+\theta)^{\alpha-1}} \, \mathrm{d}w$$

exists and is finite. Moreover, since $g(x)$ is a continuous density, there exists $M > 0$ and $0 < a < b < \infty$ such that $g(x) \geq M$ for all $x \in [a,b]$. Hence, for $\theta < a$,

$$\int_0^\infty g(w+\theta)\frac{w^{\alpha+d-1}}{(w+\theta)^{\alpha-1}} \, \mathrm{d}w \geq M \int_{a-\theta}^{b-\theta} \frac{w^{\alpha+d-1}}{(w+\theta)^{\alpha-1}} \, \mathrm{d}w > 0,$$

so $L > 0$. Thus,

$$\psi(\theta) := \theta \int_0^1 g(\theta/u) u^{-2-d}(1-u)^{\alpha+d-1}\, \mathrm{d}u \;\to\; C\theta^{-d}, \qquad\qquad \theta \to 0$$

and hence for $\delta > 0$ and θ sufficiently small, $|\psi(\theta) - C\theta^{-d}| < \delta$. Thus, for x sufficiently small,

$$\int_0^x \psi(\theta)\, \mathrm{d}\theta \leq \int_0^x C\theta^{-d}\, \mathrm{d}\theta + \int_0^x |\psi(\theta) - C\theta^{-d}|\, \mathrm{d}\theta$$
$$\leq \frac{C\, x^{1-d}}{1-d} + \delta x$$
$$\sim \frac{C\, x^{1-d}}{1-d}, \qquad x \to 0,$$

which shows that Eq. (29) holds.

$\square$

A.3 Proofs of CRM truncation bounds

A.3.1 Protobound

Lemma A.3.1 (Protobound). *Let Θ and Θ' be two discrete random measures. Let $X_{1:N}$ be a collection of random measures generated i.i.d. from Θ with $\mathrm{supp}(X_n) \subseteq \mathrm{supp}(\Theta)$, and let $Y_{1:N}$ be a collection of random variables where Y_n is generated from X_n via $Y_n \,|\, X_n \sim f(\cdot \,|\, X_n)$. Define $Z_{1:N}$ and $W_{1:N}$ analogously for Θ'. Finally, define $Q := \mathbb{1}\left[\mathrm{supp}(X_{1:N}) \subseteq \mathrm{supp}(\Theta')\right]$. If $(X_{1:N}|\Theta, \Theta', Q = 1) \overset{\mathcal{D}}{=} (Z_{1:N}|\Theta', \Theta)$ almost surely under the joint distribution of Θ, Θ', then*

$$\frac{1}{2}\|p_Y - p_W\|_1 \leq 1 - \mathbb{P}(Q = 1),$$

where p_Y, p_W are the marginal densities of $Y_{1:N}$ and $W_{1:N}$.

Proof of Lemma 4.3.1. This is the direct application of Lemma A.3.1 to CRMs, where $\Theta \sim \mathrm{CRM}(\nu)$, and Θ' is a truncation $\Theta' = \Theta_K$. The technical condition is satisfied because the weights in $X_{1:N}$ are sampled independently for each atom in Θ. $\square$

Proof of Lemma 4.4.1. This is the direct application of Lemma A.3.1 to NCRMs, where Θ is the normalization of a CRM with distribution $\mathrm{CRM}(\nu)$, and Θ' is the normalization of its truncation. The technical condition is satisfied because the conditioning on $X_{1:N} \subseteq \mathrm{supp}(\Theta')$ is equivalent to normalization of Θ'. $\square$

Proof of Lemma A.3.1. We begin by expanding the 1-norm and conditioning on both Θ and Θ' (denoted by conditioning on $\tilde{\Theta} := (\Theta, \Theta')$ for brevity):

$$\|p_Y - p_W\|_1 = \int \left| \mathbb{E}\left[\prod_{n=1}^N f(y_n|Z_n)\right] - \mathbb{E}\left[\prod_{n=1}^N f(y_n|X_n)\right] \right| \mathrm{d}y$$
$$= \int \left| \mathbb{E}\left[\mathbb{E}\left[\prod_{n=1}^N f(y_n|Z_n)|\tilde{\Theta}\right] - \mathbb{E}\left[\prod_{n=1}^N f(y_n|X_n)|\tilde{\Theta}\right]\right] \right| \mathrm{d}y.$$

Then conditioning on Q,

$$\mathbb{E}\left[\prod_{n=1}^{N} f(y_n|X_n)|\tilde{\Theta}\right] = \mathbb{E}\left[\mathbb{E}\left[\prod_{n=1}^{N} f(y_n|X_n)|\tilde{\Theta}, Q\right]|\tilde{\Theta}\right]$$

$$\stackrel{\text{a.s.}}{=} \mathbb{P}(Q=1|\tilde{\Theta})\mathbb{E}\left[\prod_{n=1}^{N} f(y_n|Z_n)|\tilde{\Theta}\right] + \mathbb{P}(Q=0|\tilde{\Theta})\mathbb{E}\left[\prod_{n=1}^{N} f(y_n|X_n)|\tilde{\Theta}, Q=0\right],$$

where the first term arises from the fact that for any function ϕ,

$$\mathbb{E}\left[\phi(Z_{1:N})|\tilde{\Theta}\right] \stackrel{\text{a.s.}}{=} \mathbb{E}\left[\phi(X_{1:N})|\tilde{\Theta}, Q=1\right],$$

because $X_{1:N}|\tilde{\Theta}, Q = 1$ is equal in distribution to $Z_{1:N}|\tilde{\Theta}$ a.s. by assumption. Substituting this back in above,

$$\|p_Y - p_W\|_1 \tag{30}$$

$$= \int\left|\mathbb{E}\left[\mathbb{P}(Q=0|\tilde{\Theta})\left(\mathbb{E}\left[\prod_{n=1}^{N} f(y_n|Z_n)|\tilde{\Theta}\right] - \mathbb{E}\left[\prod_{n=1}^{N} f(y_n|X_n)|\tilde{\Theta}, Q=0\right]\right)\right]\right|\mathrm{d}y$$

$$\leq \int\mathbb{E}\left[\mathbb{P}(Q=0|\tilde{\Theta})\left|\mathbb{E}\left[\prod_{n=1}^{N} f(y_n|Z_n)|\tilde{\Theta}\right] - \mathbb{E}\left[\prod_{n=1}^{N} f(y_n|X_n)|\tilde{\Theta}, Q=0\right]\right|\right]\mathrm{d}y$$

$$\leq \int\mathbb{E}\left[\mathbb{P}(Q=0|\tilde{\Theta})\left(\mathbb{E}\left[\prod_{n=1}^{N} f(y_n|Z_n)|\tilde{\Theta}\right] + \mathbb{E}\left[\prod_{n=1}^{N} f(y_n|X_n)|\tilde{\Theta}, Q=0\right]\right)\right]\mathrm{d}y,$$

and finally by Fubini's Theorem,

$$\|p_Y - p_W\|_1$$

$$\leq \mathbb{E}\left[\mathbb{P}(Q=0|\tilde{\Theta})\left(\mathbb{E}\left[\int\prod_{n=1}^{N} f(y_n|Z_n)\mathrm{d}y|\tilde{\Theta}\right] + \mathbb{E}\left[\int\prod_{n=1}^{N} f(y_n|X_n)\mathrm{d}y|\tilde{\Theta}, Q=0\right]\right)\right]$$

$$= \mathbb{E}\left[\mathbb{P}(Q=0|\tilde{\Theta})\left(\mathbb{E}\left[1|\tilde{\Theta}\right] + \mathbb{E}\left[1|\tilde{\Theta}, Q=0\right]\right)\right]$$

$$= 2\,\mathbb{P}(Q=0) = 2(1 - \mathbb{P}(Q=1)).$$

$\square$

Proof of Propositions 4.3.2 and 4.4.2. The same proof applies to both results. Since G is non-atomic and Ψ is Borel, there exists a measurable mapping $\mathcal{T} : \Psi \to \mathbb{R}$ such that the random variable $\mathcal{T}(\psi)$, $\psi \sim G$, is non-atomic. For an atomic measure $\varrho = \sum_{k=1}^{K_\varrho} w_k \delta_{\psi_k}$ on Ψ, define

$$S(\varrho, \mathcal{T}) := \sum_{k=1}^{K_\varrho} \mathcal{T}(\psi_k).$$

Let $\psi_i' \stackrel{\text{i.i.d.}}{\sim} G$, $T_X := \#\{k \mid x_k > 0 \wedge z_k = 0\} < \infty$, and $T_Z := \#\{k \mid z_k > 0 \wedge x_k = 0\}$. Conditional on $Q = 0$, there exists $n \in [N]$ such that $T_{X_n} \geq 1$ since X_n has at least one atom that Z_n lacks.

Using these definitions and facts, observe that

$$\mathbb{E}[S(X_n, \mathcal{T}) - S(Z_n, \mathcal{T}) \,|\, Q = 0]$$

$$= \mathbb{E}\left[\sum_{i=1}^{T_{X_n}} \mathcal{T}(\psi_i') - \sum_{j=1}^{T_{Z_n}} \mathcal{T}(\psi_j') \,|\, Q = 0\right]$$

$$= \mathbb{E}\left[\sum_{t_X=1, t_Z=0}^{\infty} \mathbb{P}(T_{X_n} = t_X, T_{Z_n} = t_Z \,|\, Q = 0)\right.$$

$$\left. \times \mathbb{E}\left[\sum_{i=1}^{t_X} \mathcal{T}(\psi_i') - \sum_{j=1}^{t_Z} \mathcal{T}(\psi_j') \,|\, T_{X_n} = t_X, T_{Z_n} = t_z, Q = 0\right]\right].$$

Since $T_{X_n} \geq 1$, $V(t_X, t_Z) := \sum_{i=1}^{t_X} \mathcal{T}(\psi_i') - \sum_{j=1}^{t_Z} \mathcal{T}(\psi_j')$ is a finite sum of one or more non-atomic random variables, so $V(t_X, t_Z)$ is itself non-atomic. It then follows that $\Delta := S(X, \mathcal{T}) - S(Z, \mathcal{T})$ is non-atomic. Thus, by Lemma A.1.2, for any δ, there exists $\epsilon_\delta > 0$ such that $\mathbb{P}(|\Delta| < \epsilon_\delta) < \delta$. Define the family of likelihoods $f_\delta(\cdot \,|\, \varrho) = \mathsf{Unif}[S(\varrho, \mathcal{T}) - \epsilon_\delta/2, S(\varrho, \mathcal{T}) + \epsilon_\delta/2]$. Let $f = f_\delta$. Then, conditioned on $Q = 0$, with probability at least $1 - \delta$, $f(y \,|\, X_n) > 0 \implies f(y \,|\, Z_n) = 0$ and $f(y \,|\, Z_n) > 0 \implies f(y \,|\, X_n) = 0$, which implies that both inequalities in Eq. (30) are equalities. Hence, we conclude that $\|p_Y - p_W\|_1 \geq 2(1 - \delta)\mathbb{P}(Q = 0)$. $\qquad\square$

A.3.2 Series representation truncation

Recall from Section 4.3.1 that a series representation generally has the form

$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k} \qquad \theta_k = \tau(V_k, \Gamma_k) \qquad V_k \overset{\text{i.i.d.}}{\sim} g,$$

where $\Gamma_k = \sum_{\ell=1}^{k} E_\ell$, $E_\ell \overset{\text{i.i.d.}}{\sim} \mathsf{Exp}(1)$, are the jumps in a unit-rate homogeneous Poisson point process, $\tau : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is a measurable function, g is a distribution on $\mathbb{R}_+$, and $\lim_{u \to \infty} \tau(v, u) = 0$ for g-almost every v. Note that by Lemma A.1.3 $\bar{\pi}(0) = 0$, where $\bar{\pi}(x) := 1 - \pi(x)$. This fact will repeatedly prove useful for the proofs in this section.

The proof of Theorem 4.3.3 is based on the following lemma.

Lemma A.3.2. *Under the same hypotheses as Theorem 4.3.3,*

$$\mathbb{P}(\mathrm{supp}(X_{1:N}) \subseteq \mathrm{supp}(\Theta_K))$$
$$= \mathbb{E}\left[\exp\left\{-\int_0^\infty \left(1 - \int_0^\infty \pi(\tau(v, u + G_K))^N g(\mathrm{d}v)\right) \mathrm{d}u\right\}\right]. \qquad (31)$$

Proof. Let

$$p(t, K) := \mathbb{E}\left[\prod_{k=K+1}^{\infty} \pi\left(\tau(V_k, \Gamma_k + t)\right)^N\right],$$

so $p(0, K) = \mathbb{P}(\text{supp}(X_{1:N}) \subseteq \text{supp}(\Theta_K))$. We use the proof strategy from Banjevic et al. [2002] and induction in K. For $K = 0$,

$$p(t, 0) = \mathbb{E}\left[\mathbb{E}\left[\pi\left(\tau(V_1, u+t)\right)^N \prod_{k=2}^{\infty} \pi\left(\tau\left(V_k, \sum_{1<j\leq k} E_j + u + t\right)\right)^N \middle| \Gamma_1 = u\right]\right]$$

$$= \int_0^{\infty} \int \pi\left(\tau(v, u+t)\right)^N p(u+t, 0) e^{-u} g(\mathrm{d}v)\, \mathrm{d}u$$

since the V_k are i.i.d. Multiplying both sides by e^{-t} and making the change of variable $w = u + t$ yields

$$e^{-t} p(t, 0) = \int_t^{\infty} \int \pi\left(\tau(v, w)\right)^N p(w, 0) e^{-w} g(\mathrm{d}v)\, \mathrm{d}w.$$

Differentiating both sides with respect to t and rearranging yields

$$\frac{\partial p(t, 0)}{\partial t} = p(t, 0)\left(1 - \int \pi\left(\tau(v, t)\right)^N g(\mathrm{d}v)\right). \tag{32}$$

Since $\lim_{u\to\infty} \tau(v, u) = 0$ and $\pi(0) = 1$ by Lemma A.1.3, we can solve Eq. (32) and conclude that

$$p(t, 0) = \exp\left\{-\int_t^{\infty}\left(1 - \int \pi\left(\tau(v, u)\right)^N g(\mathrm{d}v)\right)\mathrm{d}u\right\}$$

$$= \exp\left\{-\int_0^{\infty}\left(1 - \int \pi\left(\tau(v, u+t)\right)^N g(\mathrm{d}v)\right)\mathrm{d}u\right\}.$$

We use the inductive hypothesis that

$$p(t, K) = \mathbb{E}[p(t + G_K, 0)], \qquad G_K \sim \mathsf{Gam}(K, 1), \quad G_0 = 0,$$

which trivially holds for $K = 0$. If the inductive hypothesis holds for some $K \geq 0$, then using the tower property,

$$p(t, K+1) = \mathbb{E}\left[\mathbb{E}\left[\prod_{k=K+2}^{\infty} \pi\left(\tau\left((V_k, \sum_{1<j\leq k} E_{1j} + u + t)\right)\right)^N \middle| \Gamma_1 = u\right]\right]$$

$$= \mathbb{E}[p(t + E_1, K)], \qquad E_1 \sim \mathsf{Exp}(1)$$

$$= \mathbb{E}[p(t + G_K + E_1, 0)]$$

$$= \mathbb{E}[p(t + G_{K+1}, 0)].$$

Eq. (31) follows by setting $t = 0$. $\qquad\square$

Proof of Theorem 4.3.3. The main result follows by combining Lemmas 4.3.1 and A.3.2, applying Jensen's inequality, then using monotone convergence. The upper bound $1 - e^{-B_{N,K}} \leq 1$ follows immediately from the fact that the integral

$$\int_0^{\infty}\left(1 - \mathbb{E}\left[\int_0^{\infty} \pi(\tau(v, u+G_K))^N g(\mathrm{d}v)\right]\right)\mathrm{d}u$$

is non-negative.

Fix N. It follows from Eq. (2) that

$$\lim_{K \to \infty} 1 - \mathbb{P}(\text{supp}(X_{1:N}) \subseteq \text{supp}(\Theta_K)) \to 0.$$

It then follows from Lemma A.1.1 that $1 - e^{-\omega_{N,K}} \xrightarrow{P} 0$, where $\omega_{N,K} := \int_0^\infty \left(1 - \int_0^\infty \pi(\tau(v, u + G_K))^N g(\mathrm{d}v)\right) \mathrm{d}u$. By the continuous mapping theorem, conclude that $\omega_{N,K} \xrightarrow{P} 0$ as $K \to \infty$ and hence $B_{N,K} = \mathbb{E}[\omega_{N,K}] \to 0$ as $K \to \infty$. $\qquad\square$

Theorem A.3.3 (Inverse-Lévy representation truncation error). *For $\Theta \leftarrow \text{IL-Rep}(\nu)$, the conclusions of Theorem 4.3.3 hold with*

$$B_{N,K} = N \int_0^\infty F_K(\nu[x, \infty))(1 - \pi(x)) \, \nu(\mathrm{d}x).$$

Proof. We have from Theorem 4.3.3 that

$$B_{N,K} = N\mathbb{E}\left[\int_0^\infty (1 - \pi(\nu^{\leftarrow}(u + G_K)) \, \mathrm{d}u\right].$$

We first make the change of variables $x = \nu^{\leftarrow}(u)$ to obtain

$$\int_0^\infty \bar{\pi}(\nu^{\leftarrow}(u + G_K)) \, \mathrm{d}u = \int_{G_K}^\infty \bar{\pi}(\nu^{\leftarrow}(u)) \, \mathrm{d}u = \int_0^{\nu^{\leftarrow}(G_K)} \bar{\pi}(x)\nu(\mathrm{d}x)$$

Finally, use the fact that for all $a, b \geq 0$, $\nu^{\leftarrow}(a) \geq b \iff a \leq \nu([b, \infty))$ and monotone convergence:

$$\begin{aligned}
B_{N,K} &= N\mathbb{E}\left[\int_0^{\nu^{\leftarrow}(G_K)} \bar{\pi}(x) \, \nu(\mathrm{d}x)\right] \\
&= N\mathbb{E}\left[\int_0^\infty \mathbb{1}[x \leq \nu^{\leftarrow}(G_K)]\bar{\pi}(x) \, \nu(\mathrm{d}x)\right] \\
&= N\mathbb{E}\left[\int_0^\infty \mathbb{1}[G_K \leq \nu[x, \infty)]\bar{\pi}(x) \, \nu(\mathrm{d}x)\right] \\
&= N \int_0^\infty F_K(\nu[x, \infty))\bar{\pi}(x) \, \nu(\mathrm{d}x).
\end{aligned}$$

$\qquad\square$

Theorem A.3.4 (Thinning representation truncation error). *For $\Theta \leftarrow \text{T-Rep}(\nu)$, the conclusions of Theorem 4.3.3 hold with*

$$B_{N,K} = N \int_0^\infty (1 - \pi(v)) \int_0^{\frac{\mathrm{d}\nu}{\mathrm{d}g}(v)} F_K\left(\frac{\mathrm{d}\nu}{\mathrm{d}g}(v) - u\right) \mathrm{d}u \, g(v)\mathrm{d}v.$$

55

Proof. We have from Theorem 4.3.3 that

$$B_{N,K} = N\mathbb{E}\left[\int_0^\infty \int_0^\infty \bar{\pi}\left(v\mathbb{1}\left[\frac{\mathrm{d}\nu}{\mathrm{d}g}(v) \geq u + G_K\right]\right)\mathrm{d}u\, g(v)\mathrm{d}v\right]$$

Since $\bar{\pi}(0) = 0$, using monotone convergence we have

$$\begin{aligned}
B_{N,K} &= N\int_0^\infty \bar{\pi}(v)\int_0^\infty \mathbb{E}\left[\mathbb{1}\left[\frac{\mathrm{d}\nu}{\mathrm{d}g}(v) \geq u + G_K\right]\right]\mathrm{d}u\, g(v)\mathrm{d}v \\
&= N\int_0^\infty \bar{\pi}(v)\int_0^\infty F_K\left(\frac{\mathrm{d}\nu}{\mathrm{d}g}(v) - u\right)\mathrm{d}u\, g(v)\mathrm{d}v \\
&= N\int_0^\infty \bar{\pi}(v)\int_0^{\frac{\mathrm{d}\nu}{\mathrm{d}g}(v)} F_K\left(\frac{\mathrm{d}\nu}{\mathrm{d}g}(v) - u\right)\mathrm{d}u\, g(v)\mathrm{d}v \\
&= N\int_0^\infty \bar{\pi}(v)\int_0^{\frac{\mathrm{d}\nu}{\mathrm{d}g}(v)} F_K(u)\,\mathrm{d}u\, g(v)\mathrm{d}v.
\end{aligned}$$

$\square$

Theorem A.3.5 (Rejection representation truncation error)**.** *For $\Theta \leftarrow \mathrm{R\text{-}Rep}(\nu)$, the conclusions of Theorem 4.3.3 hold with*

$$B_{N,K} = N\int_0^\infty F_K(\mu[x,\infty))(1 - \pi(x))\,\nu(\mathrm{d}x).$$

Proof. We have from Theorem 4.3.3 that

$$B_{N,K} = N\mathbb{E}\left[\int_0^\infty \int_0^1 \bar{\pi}\left(\mu^\leftarrow(u + G_K)\mathbb{1}\left[\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\mu^\leftarrow(u + G_K)) \geq v\right]\right)\mathrm{d}v\,\mathrm{d}u\right]$$

Since $\bar{\pi}(0) = 0$, we can eliminate the innermost integral:

$$\begin{aligned}
&\int_0^1 \bar{\pi}\left(\mu^\leftarrow(u + G_K)\mathbb{1}\left[\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\mu^\leftarrow(u + G_K)) \geq v\right]\right)\mathrm{d}v \\
&= \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\mu^\leftarrow(u + G_K))\bar{\pi}(\mu^\leftarrow(u + G_K)).
\end{aligned}$$

Making the change of variable $x = \mu^\leftarrow(u + G_K)$ and reasoning analogously to the proof of Theorem A.3.3, we obtain

$$B_{N,K} = N\mathbb{E}\left[\int_0^{\mu^\leftarrow(G_K)} \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)\bar{\pi}(x)\mu(\mathrm{d}x)\right] = N\int_0^\infty F_K(\mu[x,\infty))\bar{\pi}(x)\nu(\mathrm{d}x).$$

$\square$

Theorem A.3.6 (Bondesson representation truncation error)**.** *For $\Theta \leftarrow \mathrm{R\text{-}Rep}(\nu)$, the hypotheses of Theorem 4.3.3 are satisfied and its conclusions hold with*

$$B_{N,K} = N\int_0^\infty \left(1 - \mathbb{E}\left[\pi(ve^{-G_K})\right]\right)\nu(\mathrm{d}v).$$

56

Proof. While Theorem A.3.6 can be proved using Theorem 4.3.3, we take an alternative approach using more direct Poisson process arguments.

Lemma A.3.7. *For $K \geq 0$, $G_K \sim \mathsf{Gam}(K, c)$, and $G_0 = 0$,*

$$\mathbb{P}(\mathrm{supp}(X_{1:N}) \subseteq \mathrm{supp}(\Theta_K)) = \mathbb{E}\left[\exp\left\{-\int_0^\infty \left(1 - \pi(ve^{-G_K})^N\right) \nu(\mathrm{d}v)\right\}\right].$$

Proof of Lemma A.3.7. For $t \geq 0$, the measure $t\Theta$ has distribution $t\Theta \sim \mathrm{CRM}(\nu_t)$ where $\nu_t(\mathrm{d}\theta) := \nu(\mathrm{d}\theta/t)$. Further, define $\tilde{X}_n \mid \Theta \overset{\text{i.i.d.}}{\sim} \mathrm{LP}(h, t\Theta)$, and

$$p(t, K) := \mathbb{P}(\mathrm{supp}(\tilde{X}_{1:N}) \subseteq \mathrm{supp}(t\Theta_K)) = \mathbb{E}\left[\prod_{k=K+1}^\infty \pi\left(tV_k e^{-\Gamma_k/c}\right)^N\right].$$

We will prove that

$$p(t, K) = \mathbb{E}\left[\exp\left\{-\int_0^\infty \left(1 - \pi(tve^{-G_K})^N\right) \nu(\mathrm{d}v)\right\}\right],$$

and then set $t = 1$ to obtain the desired result. The proof proceeds by induction. For $K = 0$, the event $\mathrm{supp}(\tilde{X}_{1:N}) \subseteq \mathrm{supp}(t\Theta_K)$ is equal to $\mathrm{supp}(\tilde{X}_{1:N}) \subseteq \emptyset$ and thus $\mathrm{supp}(\tilde{X}_{1:N}) = \emptyset$. This is in turn equivalent to the probability that after thinning a $\mathrm{CRM}(\nu_t)$ by $\pi(\theta)$ N times, the remaining process has no atoms, i.e. the probability that $\mathrm{CRM}\left(\left(1 - \pi(\theta)^N\right)\nu_t(\mathrm{d}\theta)\right)$ has no atoms. Since a Poisson process with measure $\mu(\mathrm{d}\theta)$ has no atoms with probability $e^{-\int \mu(\mathrm{d}\theta)}$,

$$p(t, 0) = \exp\left(-\int_0^\infty \left(1 - \pi(\theta)^N\right)\nu_t(\mathrm{d}\theta)\right)$$

$$= \mathbb{E}\left[\exp\left(-\int_0^\infty \left(1 - \pi(tve^{-G_0})^N\right)\nu(\mathrm{d}v)\right)\right].$$

The second equality follows by the change of variables $v = \theta/t$ and because $G_0 = 0$ with probability 1. The inductive hypothesis is that for $K \geq 0$,

$$p(t, K) = \mathbb{E}\left[p\left(te^{-G_K}, 0\right)\right], \qquad G_K \sim \mathsf{Gam}(K, c).$$

Using the tower property to condition on Γ_1/c and the fact that the V_k are i.i.d.,

$$p(t, K+1) = \mathbb{E}\left[\prod_{k=K+2}^\infty \pi\left(tV_k e^{-\Gamma_k/c}\right)^N\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[\prod_{k=K+1}^\infty \pi\left(te^{-\Gamma_1/c}V_k e^{-\Gamma_k/c}\right)^N \mid \Gamma_1/c\right]\right]$$

$$= \mathbb{E}\left[p\left(te^{-\Gamma_1/c}, K\right)\right] = \mathbb{E}\left[p\left(te^{-(\Gamma_1/c + G_K)}, 0\right)\right] = \mathbb{E}\left[p\left(te^{-G_{K+1}}, 0\right)\right],$$

since $\Gamma_1/c \sim \mathsf{Exp}(c)$. The desired result follows by setting $t = 1$. $\qquad\square$

First combine Lemmas 4.3.1 and A.3.7, then apply Jensen's inequality. The bounds on $B_{N,K}$ and fact that $\lim_{K\to\infty} B_{N,K} = 0$ follows by the same arguments as in the proof of Theorem 4.3.3.

$\square$

A.3.3 Superposition representation truncation

Proof of Theorem 4.3.4. We begin with Lemma 4.3.1, and note that

$$\mathbb{P}\left(\mathrm{supp}(X_{1:N}) \subseteq \mathrm{supp}(\Theta_K)\right) = \mathbb{P}\left(\mathrm{supp}(X_{1:N}) \cap \mathrm{supp}(\Theta_K^+) = \emptyset\right).$$

After generating $X_{1:N}$ from Θ, we can view the point process representing the atoms in Θ_K^+ not contained in any X_n as Θ_K^+ thinned by $\pi(\theta)^N$ (i.e. the Bernoulli trial with success probability $1 - \pi(\theta)$ to generate an atom failed N times), and thus the remaining process is Θ_K^+ thinned by $1 - \pi(\theta)^N$. Therefore, the above event is equivalent to the event that Θ_K^+ thinned by $1 - \pi(\theta)^N$ has no atoms. Using the fact that a Poisson process with measure $\mu(\mathrm{d}\theta)$ has no atoms with probability $e^{-\int \mu(\mathrm{d}\theta)}$, we have the formula for $B_{N,K}$,

$$\mathbb{P}\left(\mathrm{supp}(X_{1:N}) \subseteq \mathrm{supp}(\Theta_K)\right) = e^{-\int \left(1 - \pi(\theta)^N\right)\nu_K^+(\mathrm{d}\theta)}.$$

Since $B_{N,K} := \int \left(1 - \pi(\theta)^N\right)\nu_K^+(\mathrm{d}\theta)$ is nonnegative, the error bound lies in the interval $[0, 1]$. To show that $\lim_{K \to \infty} B_{N,K} = 0$, first note that

$$\int \left(1 - \pi(\theta)^N\right)\nu(\mathrm{d}\theta) \leq N \int (1 - \pi(\theta))\,\nu(\mathrm{d}\theta) = N \sum_{x=1}^{\infty} \int h(x \mid \theta)\nu(\mathrm{d}\theta) < \infty, \tag{33}$$

by Eq. (2). Further, splitting ν into its individual summed components, we have that

$$\int (1 - \pi(\theta))\nu(\mathrm{d}\theta) = \sum_{k=1}^{K} \int (1 - \pi(\theta))\nu_k(\mathrm{d}\theta) + \int (1 - \pi(\theta))\nu_K^+(\mathrm{d}\theta). \tag{34}$$

Combining the results from Eqs. (33) and (34) yields

$$\lim_{K \to \infty} \int (1 - \pi(\theta))\nu_K^+(\mathrm{d}\theta) = 0.$$

$\square$

A.3.4 Stochastic mapping truncation

Proof of Proposition 4.3.5. Let $\tilde{\pi}(u) = \tilde{h}(0 \mid u)$. For notational brevity, define Q to be the event where $\mathrm{supp}(X_{1:N}) \subseteq \mathrm{supp}(\Theta_K)$, and $\tilde{Q}$ to be the corresponding transformed event where $\mathrm{supp}(\tilde{X}_{1:N}) \subseteq \mathrm{supp}(\tilde{\Theta}_K)$. Then we have

$$\mathbb{P}(\tilde{Q}) = \mathbb{E}\left[\prod_{k=K+1}^{\infty} \tilde{\pi}(u_k)^{\tilde{N}}\right].$$

If $h(x \mid \theta) = \mathsf{Bern}(x; 1 - \pi_{\kappa,\tilde{N}}(\theta))$ and $N = 1$, then

$$\mathbb{P}(\tilde{Q}) = \mathbb{E}\left[\prod_{k=K+1}^{\infty} \int \tilde{\pi}(u)^{\tilde{N}} \kappa(\theta_k, \mathrm{d}u)\right] = \mathbb{P}(Q).$$

$\square$

A.3.5 Truncation with hyperpriors

Proof of Proposition 4.3.6. By repeating the proof of Lemma 4.3.1 in Appendix A.3.1, except with an additional use of the tower property to condition on the hyperparameters Φ, an additional use of Fubini's theorem to swap integration and expectation, and Jensen's inequality, we have

$$\frac{1}{2} \|p_Y - p_W\|_1 \leq \mathbb{E}\left[1 - \mathbb{P}\left(\text{supp}(X_{1:N}) \subseteq \text{supp}(\Theta_K) \mid \Phi\right)\right]$$
$$\leq \mathbb{E}\left[1 - e^{-B_{N,K}(\Phi)}\right]$$
$$\leq 1 - e^{-\mathbb{E}\left[B_{N,K}(\Phi)\right]}.$$

$\square$

A.4 Proofs of normalized truncation bounds

Proof of Lemma 4.4.3. First, we demonstrate that the $\arg\max$ is well-defined. Note that

$$\underset{i\in\mathbb{N}}{\arg\max}\, T_i + \log p_i = \underset{i\in\mathbb{N}}{\arg\max}\, \exp\left(T_i + \log p_i\right)$$

if it exists, due to the monotonicity of $\exp$. Similarly, existence of either proves the existence of the other. Since T_i are i.i.d. $\mathsf{Gumbel}(0,1)$,

$$\mathbb{P}\left(\exp\left(T_i + \log p_i\right) > \epsilon\right) = 1 - \exp\left(-e^{-(\log \epsilon - \log p_i)}\right)$$
$$= 1 - \exp\left(-p_i/\epsilon\right) \leq 1 - (1 - p_i/\epsilon).$$

Therefore,

$$\sum_{i=1}^{\infty} \mathbb{P}\left(\exp\left(T_i + \log p_i\right) > \epsilon\right) \leq \sum_{i=1}^{\infty} 1 - (1 - p_i/\epsilon) = \epsilon^{-1} \sum_{i=1}^{\infty} p_i < \infty.$$

This is sufficient to demonstrate that

$$\exp\left(T_i + \log p_i\right) \overset{a.s.}{\to} 0 \qquad \text{as } i \to \infty.$$

Finally, since any positive sequence converging to 0 can have only a finite number of elements greater than any $\epsilon > 0$, set $\epsilon = \exp(T_1 + \log p_1)$, and thus

$$\underset{i\in\mathbb{N}}{\arg\max}\, \exp\left(T_i + \log p_i\right) = \underset{i:T_i+\log p_i \geq \epsilon}{\arg\max}\, \exp\left(T_i + \log p_i\right)$$

where the right hand side exists because it computes the maximum of a finite, nonempty set of numbers. Note that the $\arg\max$ is guaranteed to be a single element, since $T_i + \log p_i$ has a purely diffuse distribution on $\mathbb{R}$.

Now that the a.s. existence and uniqueness of the $\arg\max$ has been demonstrated, we can compute its distribution. First, note that

$$\mathbb{P}\left(T_i + \log p_i \leq x \,\forall i \in \mathbb{N}, i \neq j\right) = \prod_{i=1,i\neq j}^{\infty} \exp\left(-e^{-(x-\log p_i)}\right) = \exp\left(-e^{-x}(s - p_j)\right),$$

where $s := \sum_i p_i$. So then

$$\mathbb{P}\left(j = \arg\max_{i \in \mathbb{N}} T_i + \log p_i \right) = \mathbb{P}\left(T_i + \log p_i \leq T_j + \log p_j \; \forall i \in \mathbb{N} \right)$$

$$= \int e^{-e^{-x}(s-p_j)} e^{-\left(x - \log p_j + e^{-(x - \log p_j)} \right)} \mathrm{d}x$$

$$= p_j \int e^{-se^{-x}} e^{-x} \mathrm{d}x$$

$$= \frac{p_j}{\sum_i p_i} \int e^{-e^{-(x - \log s)}} e^{-(x - \log s)} \mathrm{d}x$$

$$= \frac{p_j}{\sum_i p_i},$$

where the last integral is 1 since its integrand is the Gumbel$(s, 1)$ density. $\qquad\square$

A.4.1 Normalized series representation truncation

Proof of Theorem 4.4.4. First, we apply Lemma 4.4.1,

$$\frac{1}{2} \| p_Y - p_W \|_1 \leq 1 - \mathbb{P}\left(X_{1:N} \subseteq \operatorname{supp}(\Xi_K) \right).$$

Next, by Jensen's inequality,

$$\mathbb{P}\left(X_{1:N} \subseteq \operatorname{supp}(\Xi_K) \right) = \mathbb{E}\left[\left(\frac{\Theta_K(\Psi)}{\Theta(\Psi)} \right)^N \right] \geq \mathbb{E}\left[\frac{\Theta_K(\Psi)}{\Theta(\Psi)} \right]^N$$

$$= \mathbb{P}(X_1 \in \operatorname{supp}(\Xi_K))^N.$$

The remaining part of this proof quantifies the probability that sampling X_1 from Ξ generates an atom in the support of Ξ_K (equal to the support of Θ_K, since Ξ is just the normalization of Θ). To do this, we use the trick based on Lemma 4.4.3: we log-transform the rates in Θ, perturb them all by i.i.d. Gumbel random variables, and quantify the probability that the max occurs within the atoms of Θ_K.

First, we split the sequential representation of Θ into the truncation Θ_K and its tail Θ_K^+, using the form from Eq. (6),

$$\Theta = \sum_{k=1}^{\infty} \tau(V_k, \Gamma_k) \delta_{\psi_k} = \sum_{k=1}^{K} \tau(V_k, \Gamma_k) \delta_{\psi_k} + \sum_{k=K+1}^{\infty} \tau(V_k, \Gamma_k) \delta_{\psi_k} := \Theta_K + \Theta_K^+.$$

Next, we define the maximum of the log-transformed, Gumbel perturbed rates in Θ_K as

$$M_K := \max_{1 \leq k \leq K} \log \tau(V_k, \Gamma_k) + W_k,$$

where $W_k \overset{\text{i.i.d.}}{\sim} \operatorname{Gumbel}(0, 1)$. Since Γ_k are from a unit-rate homogeneous Poisson process, if we condition on the value of Γ_K, this is equivalent to conditioning on the event that the Poisson process

60

has exactly $K - 1$ atoms on $[0, \Gamma_K]$, and the Kth atom is at Γ_K. And since M_K does not depend on the ordering of $(\Gamma_k)_{k=1}^K$,

$$M_K \,|\, \Gamma_K \stackrel{\mathcal{D}}{=} \max \left\{ \log \tau(V_K, \Gamma_K) + W_K, \max_{1 \leq k \leq K-1} \log \tau(V_k, U_k) + W_k \right\},$$

where $U_k \stackrel{\text{i.i.d.}}{\sim} \text{Unif}(0, \Gamma_K)$. Therefore, $M_K \,|\, \Gamma_K$ is the maximum of a collection of independent random variables, so we can compute its CDF by simply taking the product of the CDFs of each of those random variables. Using standard techniques for transformation of independent random variables, we have that

$$\mathbb{P}\left(\log \tau(V_k, U_k) + W_k \leq x \,|\, \Gamma_K \right) = \int_0^\infty \int_0^1 g(v) e^{-\tau(v, \Gamma_K u) e^{-x}} \mathrm{d}u \mathrm{d}v, \quad k < K$$

$$\mathbb{P}\left(\log \tau(V_K, \Gamma_K) + W_K \leq x \,|\, \Gamma_K \right) = \int_0^\infty g(v) e^{-\tau(v, \Gamma_K) e^{-x}} \mathrm{d}v,$$

so

$$\mathbb{P}\left(M_K \leq x \,|\, \Gamma_K \right) = \left(\int_0^\infty \int_0^1 g(v) e^{-\tau(v, \Gamma_K u) e^{-x}} \mathrm{d}u \mathrm{d}v \right)^{K-1} \left(\int_0^\infty g(v) e^{-\tau(v, \Gamma_K) e^{-x}} \mathrm{d}v \right).$$

Defining the function

$$J(u, t) := \mathbb{E}\left[e^{-t \cdot \tau(V, u)} \right], \qquad V \sim g,$$

we have

$$\mathbb{P}\left(M_K \leq x \,|\, \Gamma_K \right) = \left(\int_0^1 J\left(\Gamma_K u, e^{-x} \right) \mathrm{d}u \right)^{K-1} J\left(\Gamma_K, e^{-x} \right).$$

Next, we define an analogous maximum for the tail process rates in Θ_K^+,

$$M_K^+ := \sup_{k > K} \log \tau(V_k, \Gamma_k) + W_k.$$

Conditioning on Γ_K, we have

$$M_K^+ \,|\, \Gamma_K \stackrel{\mathcal{D}}{=} \sup_{k \geq 1} \log \tau(V_k, \Gamma_k' + \Gamma_K) + W_k,$$

where Γ_k' is a unit-rate homogeneous Poisson process on $\mathbb{R}_+$. Now note that since Γ_k' is a Poisson point process on $\mathbb{R}_+$, so is $\log \tau(V_k, \Gamma_k' + \Gamma_K) + W_k$ (using Poisson process stochastic mapping), with rate measure

$$\left(\int_0^\infty \int_0^\infty e^{-(t - \log \tau(v, u + \Gamma_K)) - e^{-(t - \log \tau(v, u + \Gamma_K))}} g(v) \mathrm{d}u \mathrm{d}v \right) \mathrm{d}t.$$

Therefore, $\mathbb{P}\left(M_K^+ \leq x \,|\, \Gamma_K \right)$ is equal to the probability that the above Poisson point process has no atoms with position greater than x. Since a Poisson process on $\mathbb{R}_+$ with measure μ has no atoms above x with probability $e^{-\int_x^\infty \mu(\mathrm{d}t)}$,

$$\mathbb{P}\left(M_K^+ \leq x \,|\, \Gamma_K \right) = e^{-\int_x^\infty \left(\int_0^\infty \int_0^\infty e^{-(t - \log \tau(v, u + \Gamma_K)) - e^{-(t - \log \tau(v, u + \Gamma_K))}} g(v) \mathrm{d}u \mathrm{d}v \right) \mathrm{d}t}.$$

Noticing that the integrand in t is a Gumbel density, we can use Fubini's theorem to swap integrals and evaluate:

$$\mathbb{P}\left(M_K^+ \leq x \mid \Gamma_K\right) = e^{\int_0^\infty \int_0^\infty \left(e^{-\tau(v, u+\Gamma_K)e^{-x}} - 1\right)g(v)\mathrm{d}u\mathrm{d}v}$$

$$= e^{\int_0^\infty \left(J\left(u+\Gamma_K, e^{-x}\right) - 1\right)\mathrm{d}u}.$$

Taking the derivative yields the density of $M_K^+ \mid \Gamma_K$ with respect to the Lebesgue measure. Therefore, using the Gumbel-max trick from Lemma 4.4.3, we substitute the results for $M_K \mid \Gamma_K$ and $M_K^+ \mid \Gamma_K$ into the original bound yielding

$$\frac{1}{2}\left\|p_Y - p_W\right\|_1 \leq 1 - \mathbb{P}\left(X_1 \subseteq \mathrm{supp}(\Xi_K)\right)^N$$

$$= 1 - \left(1 - \mathbb{E}\left[\mathbb{P}\left(M_K < M_K^+ \mid \Gamma_K\right)\right]\right)^N,$$

where (using the substitution $t = e^{-x}$)

$$\mathbb{P}\left(M_K < M_K^+ \mid \Gamma_K\right) = \int_{-\infty}^\infty \mathbb{P}\left(M_K \leq x \mid \Gamma_K\right)\frac{\mathrm{d}}{\mathrm{d}x}\mathbb{P}\left(M_K^+ \leq x \mid \Gamma_K\right)\mathrm{d}x$$

$$= \int_0^\infty J\left(\Gamma_K, t\right)\left(\int_0^1 J\left(\Gamma_K u, t\right)\mathrm{d}u\right)^{K-1}\left(-\frac{\mathrm{d}}{\mathrm{d}t}e^{\int_0^\infty (J(u+\Gamma_K, t)-1)\mathrm{d}u}\right)\mathrm{d}t.$$

The fact that the bound is between 0 and 1 is a simple consequence of the fact that $\mathbb{P}\left(X_1 \in \mathrm{supp}(\Xi_K)\right) \in [0, 1]$, and the asymptotic result follows from the fact that $\mathbb{P}\left(X_1 \in \mathrm{supp}\left(\Xi_K\right)\right) \to 1$ as $K \to \infty$. $\qquad\square$

A.4.2 Normalized superposition representation truncation

Proof of Theorem 4.4.5. The same initial technique as in the proof of Theorem 4.4.5 yields

$$\frac{1}{2}\left\|p_Y - p_W\right\|_1 \leq 1 - \mathbb{P}\left(X_1 \in \mathrm{supp}\left(\Xi_K\right)\right)^N.$$

The remaining part of this proof quantifies the probability that sampling X_1 from Ξ generates an atom in the support of Ξ_K. Since most of the following developments are similar for Θ_K, ν_K and Θ_K^+, ν_K^+, we will focus the discussion on Θ_K, ν_K and reintroduce the tail quantities when necessary. First, we transform the rates of Θ_K under the stochastic mapping $w = \log\theta + W$, where $W \overset{\text{i.i.d.}}{\sim} \mathrm{Gumbel}(0, 1)$, resulting in a new Poisson point process with rate measure

$$\left(\int e^{-(t-w)-e^{-(t-w)}}e^w\nu_K(e^w)\mathrm{d}w\right)\mathrm{d}t.$$

The probability that all points in this Poisson point process are less than a value x is equal to the probability that there are no atoms above x. Defining M_K to be the supremum of the points in this process, combined with the basic properties of Poisson point processes, we have

$$\mathbb{P}(M_K \leq x) = \exp\left(-\int_x^\infty \int e^{-(t-w)-e^{-(t-w)}}e^w\nu_K(e^w)\mathrm{d}w\mathrm{d}t\right)$$

Using Fubini's theorem to swap the integrals, we can evaluate the inner integral analytically by noting the integrand is a $\mathsf{Gumbel}(w, 1)$ density,

$$
\begin{aligned}
\mathbb{P}(M_K \leq x) &= \exp\left(-\int (1 - e^{-e^{-(x-w)}}) e^w \nu_K(e^w) \mathrm{d}w\right) \\
&= \exp\left(\int \left(e^{-\theta e^{-x}} - 1\right) \nu_K(\mathrm{d}\theta)\right).
\end{aligned}
$$

We can take the derivative with respect to x to obtain its density with respect to the Lebesgue measure. The above derivation holds true for Θ_K^+, replacing ν_K with ν_K^+ and M_K with M_K^+. Therefore, using the Gumbel-max trick from Lemma 4.4.3, we substitute the results for M_K and M_K^+ into the original bound, yielding

$$
\begin{aligned}
\frac{1}{2}\|p_Y - p_W\|_1 &\leq 1 - \mathbb{P}\left(X_1 \in \mathrm{supp}\left(\Theta_K\right)\right)^N \\
&= 1 - \left(1 - \mathbb{P}\left(M_K < M_K^+\right)\right)^N,
\end{aligned}
$$

where (using the substitution $t = e^{-x}$, and the fact that $\theta e^{-\theta t}$ is dominated by θ to swap integration and differentiation)

$$
\begin{aligned}
\mathbb{P}\left(M_K < M_K^+\right) &= \int_{-\infty}^{\infty} \mathbb{P}(M_K \leq x) \frac{\mathrm{d}}{\mathrm{d}x} \mathbb{P}\left(M_K^+ \leq x\right) \mathrm{d}x \\
&= \int_{-\infty}^{\infty} e^{\int \left(e^{-\theta e^{-x}} - 1\right)\nu(\mathrm{d}\theta)} \frac{\mathrm{d}}{\mathrm{d}x}\left(\int \left(e^{-\theta e^{-x}} - 1\right) \nu_K^+(\mathrm{d}\theta)\right) \mathrm{d}x \\
&= -\int_0^{\infty} e^{\int \left(e^{-\theta t} - 1\right)\nu(\mathrm{d}\theta)} \frac{\mathrm{d}}{\mathrm{d}t}\left(\int \left(e^{-\theta t} - 1\right) \nu_K^+(\mathrm{d}\theta)\right) \mathrm{d}t \\
&= \int_0^{\infty} e^{\int \left(e^{-\theta t} - 1\right)\nu(\mathrm{d}\theta)} \left(\int \theta e^{-\theta t} \nu_K^+(\mathrm{d}\theta)\right) \mathrm{d}t.
\end{aligned}
$$

The fact that the bound lies between 0 and 1 is a simple consequence of the fact that $\mathbb{P}\left(X_1 \in \mathrm{supp}(\Xi_K)\right) \in [0, 1]$. The fact that the error bound asymptotically approaches 0 is a consequence of the monotone convergence theorem applied to the decreasing sequence of functions $\theta \nu_K^+(\mathrm{d}\theta)$. $\qquad\square$

A.4.3 Truncation with hyperpriors

Proof of Proposition 4.4.6. By repeating the proof of Lemma 4.4.1 in Appendix A.3.1, except with an additional use of the tower property to condition on the hyperparameters Φ, an additional use of Fubini's theorem to swap integration and expectation, and Jensen's inequality, we have

$$
\begin{aligned}
\frac{1}{2}\|p_Y - p_W\|_1 &\leq \mathbb{E}\left[1 - \mathbb{P}\left(X_{1:N} \subseteq \mathrm{supp}(\Theta_K) \mid \Phi\right)\right] \\
&\leq 1 - \mathbb{E}\left[(1 - B_K(\Phi))^N\right] \\
&\leq 1 - (1 - \mathbb{E}\left[B_K(\Phi)\right])^N.
\end{aligned}
$$

$\qquad\square$

List of Acronyms

Acronym	Full Phrase
BNP	Bayesian nonparametrics
BP	Beta process
BPP	Beta prime process
B-rep	Bondesson representation
CDF	Cumulative density function
CRM	Completely random measure
DB-Rep	Decoupled Bondesson representation
DP	Dirichlet process
ΓP	Gamma process
IL-Rep	Inverse-Lévy representation
LP	Likelihood process
LomP	Lomax process
MCMC	Markov Chain Monte Carlo
NCRM	Normalized completely random measure
NΓP	Normalized gamma process
PL-rep	Power-law representation
R-rep	Rejection representation
SB-rep	size-biased representation
T-rep	Thinning representation
VB	Variational Bayes